

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Deep Alignment and Super-Resolution of Computed Tomography Coronary Angiography Images

Margarida de Oliveira Fontes Pereira



Master in Biomedical Engineering

Supervisor: João Manuel Pedrosa PhD

July 18, 2024

Deep Alignment and Super-Resolution of Computed Tomography Coronary Angiography Images

Margarida de Oliveira Fontes Pereira

Master in Biomedical Engineering

July 18, 2024

Abstract

Coronary artery disease (CAD) is the leading global cause of mortality, characterized by plaque buildup in coronary arteries that restricts blood flow and can lead to myocardial infarction. Current diagnostic methods begin with initial Computed Tomography (CT) scans to detect calcifications, followed by higher-resolution CT Coronary Angiography (CTCA) with contrast to visualize blood flow and identify stenosis. Combining these scans could provide additional insights, but is hindered by slice misalignments due to patient motion between exams and differing axial resolutions, with CT scans having a 1.5 mm slice thickness and CTCA ranging from 0.4 mm to 1 mm. Therefore, developing frameworks for Super Resolution (SR) and alignment of CT and CTCA images is crucial for advancing cardiovascular disease diagnosis. While deep learning (DL) techniques like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have shown success in SR, there remains a research gap specifically in axial SR applied to cardiac CT, which is the focus of this dissertation.

This investigation into SR reconstruction in cardiac CT imaging explored two main strategies, which were tested on a private dataset obtained at CHVNGE composed of CT and CTCA scans from 203 CAD-screened patients. The first approach focused on leveraging solely CT data to intermediate reconstruction, validating SR methods by comparing them directly to original full-resolution slices. The U-Net model, particularly with MSE loss, demonstrated superior performance with an average Structural Similarity Index Measure (SSIM) of 0.994 ± 0.0064 , Peak Signal-to-Noise Ratio (PSNR) of 42.787 ± 4.52 , and Fréchet Inception Distance (FID) value of 6.532.

The second approach integrated CTCA data to generate super-resolved CT reconstructions, including previously unavailable slices at CTCA-derived positions. Prior to integration, precise spatial alignment between CT and CTCA scans was ensured using an alignment module. These methods enhances CT scan axial resolution to match CTCA data, achieving high-quality reconstructions with exceptional detail. The U-Net model, similar to the first approach, delivered superior results with a Dice Similarity Coefficient (DSC) of 0.719 ± 0.121 and FID value of 6.8898, highlighting its performance and potential for broader CT imaging applications. However, integrating CTCA data into Autoencoders (AEs) and Connected Autoencoders (CAEs) to enhance CT reconstruction increased method complexity and introduced challenges in achieving high-fidelity reconstructions. Though DSC improved to 0.749 ± 0.125 with CTCA information, indicating potential performance enhancement, realism decreased with observed misalignments in reconstructed CT images, as evidenced by higher FID values (84.4014).

In conclusion, this research successfully advanced cardiac CT imaging and CAD diagnosis, achieving higher axial SR in CT scans. These methods have potential to significantly improve diagnostic accuracy and efficiency in clinical practice. Future improvements, such as integrating dynamic alignment modules, will further enhance these methods' capabilities.

Keywords: Alignment, Computed Tomography, Computed Tomography Coronary Angiography, Coronary artery disease, Deep Learning, Super Resolution

Resumo

A Doença da Artéria Coronária (DAC) é a principal causa global de mortalidade, caracterizada pela acumulação de placas nas artérias coronárias, o que restringe o fluxo sanguíneo para o coração e pode resultar em enfartes do miocárdio. Os métodos diagnósticos atuais começam com Tomografias Computadorizadas (TC) para detectar calcificações, seguidas de Angiografias Coronárias por TC (ACTC) com contraste para visualizar o fluxo sanguíneo e identificar estenoses. Assim, a combinação de ambas as modalidades de imagem médica pode fornecer informações adicionais, mas é impossibilitada por desalinhamentos entre fatias das imagens devido ao movimento do paciente entre os exames e por diferentes resoluções axiais, com TC com espessura de corte de 1,5 mm e ACTC variando de 0,4 mm a 1 mm.

Portanto, desenvolver *frameworks* para alcançar a Super-Resolução (SR) e alinhamento entre as imagens de TC e ACTC é crucial para o avanço do diagnóstico cardiovascular. Embora técnicas de *Deep Learning* (DL), como Redes Neurais Convolucionais (CNNs) e Redes Adversárias Generativas (GANs), tenham demonstrado sucesso em SR, ainda existe uma lacuna na literatura especificamente em SR axial aplicada à TC cardíaca, que é o foco desta dissertação.

A presente pesquisa investigou duas principais estratégias de reconstrução por SR em imagens de TC cardíaca, que foram testadas num conjunto de dados privado proveniente do CHVNGE e composto por exames de TC e ACTC de 203 pacientes rastreados para DAC. A primeira estratégia focou-se exclusivamente em dados de TC para reconstruir fatias intermédias de TC, validando diretamente os métodos de SR com comparações às fatias originais de resolução completa. O modelo *U-Net*, utilizando a função de perda *MSE*, obteve um desempenho superior em comparação ao método de interpolação, alcançando um Índice de Similaridade Estrutural (*SSIM*) médio de $0,994 \pm 0,0064$, Relação Sinal-Ruído de Pico (*PSNR*) de $42,787 \pm 4,52$ e um valor de Distância *Fréchet Inception* (*FID*) de 6,532.

Na segunda abordagem, foram integrados dados de ACTC para gerar reconstruções super-resolvidas de TC, incluindo fatias previamente indisponíveis, nomeadamente nas posições correspondentes de ACTC. Para isso, antes da integração dos dados de ACTC nos métodos desenvolvidos, foi necessário desenvolver um módulo de alinhamento espacial. Essas técnicas aumentaram a resolução axial dos dados de TC para corresponder à resolução dos dados de ACTC, resultando em reconstruções de alta qualidade. À semelhança da abordagem anterior, o modelo *U-Net* demonstrou um desempenho superior, com um Coeficiente de Similaridade de Dice (*DSC*) de $0,719 \pm 0,121$ e um valor *FID* de 6,8898, destacando o seu potencial para aplicações de imagem de TC.

Contudo, a integração de dados de ACTC nos métodos de *Autoencoders* (AEs) e *Autoencoders Conectados* (CAEs), com o objetivo de melhorar a reconstrução de TC, aumentou a complexidade dos métodos, introduzindo novos desafios, em específico desalinhamentos que comprometeram o realismo das imagens reconstruídas. Embora o *DSC* tenha melhorado para $0,749 \pm 0,125$ com a incorporação de informações da ACTC, observaram-se desalinhamentos e um aumento significativo no valor *FID* (84,4014), indicando uma perda considerável de realismo.

Em suma, esta pesquisa apresenta um contributo relevante para o estado da arte no diagnóstico

por imagem de TC cardíaca e DAC, especialmente na aplicação de SR axial. Os métodos desenvolvidos têm potencial para substancialmente melhorar a precisão e eficiência do diagnóstico, beneficiando diretamente a prática clínica. No entanto, melhorias futuras, como o refinamento do método de alinhamento, são necessárias para aprimorar os métodos apresentados.

Palavras-chave: Alinhamento, Angiografia Coronária por Tomografia Computadorizada, *Deep Learning*, Doença da Artéria Coronária, Super Resolução, Tomografia Computadorizada.

Agradecimentos

Primeiramente e mais importante, é impossível não expressar a minha imensa gratidão e admiração pelo meu orientador, Professor Doutor João Pedrosa, pela sua disponibilidade ímpar, competência extraordinária, atenção e apoio incessantes, que foram fundamentais durante todo o processo desta dissertação. A sua contribuição foi inestimável e fez totalmente a diferença não só na elaboração deste trabalho, mas também servindo como um pilar no meu crescimento académico, que levo como um grande exemplo para o início da minha futura vida profissional. Além disso, sou grata pelo privilégio que foi integrar o grupo de investigação Bioimaging do INESC TEC, que me acolheu calorosamente. Ter a oportunidade de acompanhar e aprender de perto com profissionais que contribuem de forma inspiradora para a Ciência foi, sem dúvida, uma experiência enriquecedora.

Aos meus pais e ao meu irmão, não há forma de conseguir agradecer por tanto! Toda a minha vida foram e todo o resto da minha vida serão os meus maiores pilares e fontes de força que me impulsionam a alcançar os meus objetivos. Agradeço profundamente pelo apoio incondicional em todas as situações, pela preocupação diária, pelo carinho que nunca falta e pela educação excecional que me inculcaram e que me possibilitaram receber. São o lar para o qual quero sempre voltar. Agradeço também ao restante da minha família fantástica, cujo apoio e votos constantes pelo meu sucesso são uma fonte inesgotável de inspiração.

Ao Diogo, o meu companheiro nesta que é a melhor das viagens, que acompanhou de perto o meu percurso escolar e académico, sem nunca deixar de me motivar, de me fazer acreditar no meu potencial e de se prontificar para ajudar em tudo o que está ao seu alcance. Agradeço imensamente pelo companheirismo, pela paciência incansável, e por muito mais.

Por último, mas não menos importante, quero agradecer ao meu círculo de amigos, que é como uma segunda família, por todos os momentos, experiências e memórias que temos colecionado e que continuaremos a colecionar daqui em diante.

“A rock pile ceases to be a rock pile the moment a single man contemplates it, bearing within him the image of a cathedral.”

Antoine de Saint-Exupéry

Contents

1	Introduction	1
1.1	Context	1
1.2	Goals	2
1.3	Document Structure	2
2	Background	3
2.1	Heart Anatomy	3
2.2	Coronary Artery Disease	5
2.2.1	Etiology	5
2.2.2	Epidemiology	6
2.2.3	Diagnosis	6
2.3	Joint CT/CTCA Analysis	9
3	Automatic Super Resolution of CT Images	13
3.1	State of Art	14
3.1.1	CNNs	16
3.1.2	GANs	21
3.2	Evaluation Metrics	24
3.2.1	Paired Metrics	24
3.2.2	Unpaired Metrics	27
3.2.3	Visual Assessment	29
3.3	Summary	29
4	Intermediate CT Reconstruction	31
4.1	Data	32
4.2	Methodology	33
4.2.1	Interpolation-Based CT Reconstruction	33
4.2.2	CNN-Based CT Reconstruction	34
4.3	Evaluation	37
4.4	Results and Discussion	38
4.5	Summary	44
5	CT Reconstruction with CTCA Integration	45
5.1	Methodology	46
5.1.1	Data	46
5.1.2	Interpolation-Based CT Reconstruction	47
5.1.3	CNN-Based CT Reconstruction	49
5.1.4	Autoencoder with CTCA information	50

5.1.5	2D Connected Autoencoders	52
5.1.6	3D Connected Autoencoders	55
5.2	Evaluation	62
5.3	Results and Discussion	64
5.3.1	CT Reconstruction	64
5.3.2	Autoencoder with CTCA information	67
5.3.3	Connected Auto-Encoders	72
6	Conclusions and Future Work	79
	References	81

List of Figures

2.1	(a) Blood flow through the heart; (b) The heart inside the whole circulatory system.	4
2.2	Cross Section of a Coronary Artery.	4
2.3	Illustration of the formation of the atherosclerotic plaque in CAD.	5
2.4	CACS exam without (left) and with (right) calcium in the arteries.	8
2.5	CAD-RADS Reporting and Data System for patients presenting with stable chest pain.	9
2.6	CTCA exams with the visualization of stenosis in the coronary arteries grading from 0–25% to 100% occlusions.	10
3.1	Network structure of 3DSRCNN contains 12 layers 3D-CNN.	16
3.2	Diagram of model training for a specific plane.	17
3.3	Architecture of the slice-interpolation network.	18
3.4	CNN architecture with data flow.	19
3.5	a) Marginal Super-Resolution Pipeline. b) Two-view Fusion Pipeline. c) Refinement Pipeline.	20
3.6	The overall pipeline of Spatially Aware Interpolation NeTwork (SAINT). For visualization purpose, the volumes are rendered in 3D based on their bone structure.	21
3.7	Adversarial training framework of thick-thin slice translation on CT images. In each training iteration, the thin slice input data is randomly degraded to simulate thick slice data. The generator input is a 1 mm spline interpolated 3D thick slice image itself and the discriminator input is the generated Virtual Thin image with slice condition including body part information with degraded parameter scales.	22
3.8	The overview of our model. An encoder is trained to produce low-level feature maps, which are interpolated with interpolation coefficient α and the decoder tries to disentangle the latent representation and map it into data space. D1 is the discriminator trying to predict α , and D2 is trained to distinguish whether the input is real or not. (b) Network architecture of the autoencoder part. "EN" or "DE" stands for encoder or decoder block.	23
3.9	Layout of the deep learning network.	24
4.1	Schematic diagram illustrating the process of SR reconstruction using interleaved CT data.	32
4.2	Slice of CT exam from the CHVNGE dataset.	33
4.3	Corresponding slice of CTCA exam.	33
4.4	Scheme of the Interpolation-Based CT Reconstruction method.	34
4.5	General outline of the CNN-Based CT Reconstruction model training goal.	34
4.6	U-Net architecture used.	35

4.7	Result example of CT reconstruction using the interpolation method described. Input slices at the top, real intermediate slice at the bottom left and interpolated slice at the bottom right.	41
4.8	Result of model CT reconstruction using the MSE loss function: Input slices at the top, real intermediate slice at the bottom left and reconstructed slice at the bottom right.	42
4.9	Results of reconstructing the intermediate slice using the interpolation method (middle) and the U-Net model using the MSE loss function (right), in comparison to the real target slice.	43
5.1	General Schematic Diagram representing the objective of the SR CT Reconstruction.	46
5.2	Scheme of the alignment process between CT and CTCA.	47
5.3	Visual result of the alignment step.	48
5.4	Scheme of the Interpolation-Based CT Reconstruction.	49
5.5	Strategy for reconstructing CT using the U-Net model.	49
5.6	Model for enhanced CT slice reconstruction using auxiliary information from CTCA slices.	50
5.7	Hounsfield scale with simplified reference values for different structures.	51
5.8	General structure of the CAEs model.	52
5.9	General architectural framework of the 2D connected AEs.	54
5.10	General architectural framework of the 3D connected AEs.	56
5.11	Illustrative example of how the implemented batch sampler works.	57
5.12	General architectural framework of: a) CT autoencoder and b) CTCA autoencoder.	58
5.13	Diagram depicting the sampling process operation.	61
5.14	Result example of the application of the threshold, for computing the DSC. . . .	63
5.15	Result example of slice reconstruction for the Interpolation-Based method. . . .	66
5.16	Result example of slice reconstruction for the CNN-Based method.	66
5.17	Result example of slice reconstruction for the Interpolation-Based method, along with the corresponding CTCA slice.	66
5.18	Result example of slice reconstruction for the CNN-Based method, along with the corresponding CTCA slice.	67
5.19	Results examples of slice reconstruction using the Interpolation-Based method (middle) and the CNN-Based method (right), along with the corresponding CTCA slice(left).	68
5.20	Results examples of slice reconstruction for AE with CTCA information, using as CT inputs the results from the interpolation-based method.	70
5.21	Results examples of slice reconstruction for AE with CTCA information, using as CT inputs the results from the CNN-based method.	71
5.22	Results examples of slice reconstruction for the 2D CAEs.	75
5.23	Result example of slice reconstruction for the 3D CAEs. The first column shows-cases the two input CT slices, followed by the second column depicting the N input CTCA slices located between the CT input slices. The third column displays the reconstructed CT images, the fourth column shows the reconstructed CTCA images, and the final column illustrates the reconstructed contrast. Each line corresponds to slices from 1 to N.	77

5.24 Result example of slice reconstruction for the 3D CAEs. The first row showcases the two input CT slices, followed by the second row depicting the N input CTCA slices located between the CT input slices. The third row displays the reconstructed CT images, the fourth row shows the reconstructed CTCA images, and the final row illustrates the reconstructed contrast. Each line corresponds to slices from 1 to N. 78

List of Tables

3.1	Summary of Deep Learning algorithms for super resolution.	15
4.1	Comparison of Interpolation and U-Net with different loss functions results. . . .	38
5.1	Results for CT Reconstruction methods (Interpolation VS CNN-Based).	64
5.2	Results of the CT Reconstruction methods VS AE with CTCA information methods.	69
5.3	Summarization of the results of the different methods.	73

Abbreviations and Symbols

2D	Two-Dimensional
3D	Three-Dimensional
AE	Autoencoder
AI	Artificial Intelligence
CACS	Coronary Artery Calcium Score
CAD	Coronary Artery Disease
CAD-RADS	Coronary Artery Disease Reporting and Data System
CAEs	Connected Autoencoders
CHVNGE	Centro Hospitalar de Vila Nova de Gaia e Espinho
CNN	Convolutional Neural Networks
CTCA	Computed Tomography Coronary Angiography
DL	Deep Learning
DSC	Dice Similarity Coefficient
FID	Fréchet Inception Distance
FOV	Field of View
GANs	Generative Adversarial Networks
HR	High-Resolution
HU	Hounsfield Unit
ICCM	Iodine Containing Contrast Medium
LR	Low-Resolution
MSE	Mean Squared Error
NCCT	Non-Contrast Computed Tomography
PSNR	Peak Signal-to-Noise Ratio
SR	Super Resolution
SSIM	Structural Similarity Index Measure

Chapter 1

Introduction

1.1 Context

Coronary artery disease (CAD) stands as the leading cause of death worldwide, responsible for over 9 million annual fatalities globally and 7,100 in Portugal, and it is anticipated to maintain this unsettling status in the foreseeable future. It is characterized by the accumulation of calcium, fatty lipids, and inflammatory cells within the coronary arteries (atherosclerosis) and poses a severe threat to public health [1]. It impedes the normal blood flow through the coronary arteries to the heart, resulting in ischemia of the cardiac muscle cells and ultimately leading to myocardial infarction, or more commonly known, heart attack [2].

The diagnostic journey for CAD conventionally entails a two-step sequential imaging approach. Initially, non-contrast CT scans assess coronary artery calcium score (CACS), which identifies calcifications, indicative of atherosclerosis, that serve as prognostic indicators for the risk of CAD [3]. This initial assessment is subsequently followed by a higher-resolution CT coronary angiography (CTCA) examination [4] that employs contrast agents to accentuate the blood flow through the coronary arteries, enabling the detection of non-calcified stenosis regions.

Non-Contrast CT (NCCT) offers valuable diagnostic information that is currently not fully utilized in clinical practice [5]. Although CTCA has exceptional sensitivity in the detection and grading of CAD, it requires contrast agents, increasing risk and additional radiation exposure for subjects. Enhancing NCCT's diagnostic potential with advanced image analysis tools could potentially rule out CAD in a significant number of patients, allowing them to avoid CTCA, thereby saving time, resources, and reducing unnecessary radiation exposure [6].

Given the profound impact of CAD on global morbidity and mortality rates, automatic solutions to enhance the accuracy and efficiency of CAD diagnosis have been explored in recent years, focusing either on the CT or CTCA exams. Noteworthy examples include CNN-based approaches for coronary artery calcium score (CACS) [7], which aid in CT analysis, as well as the implementation of 3D CNNs for coronary artery lumen segmentation in CTCA for the quantification of stenosis [8].

Yet, an unexplored potential lies in integrating information from both modalities, harnessing the potential synergy between their respective datasets. The combination of information from these two modalities holds promise in providing a more comprehensive understanding of CAD, potentially leading to more precise diagnostic tools and improved patient outcomes. However, this approach has been impeded by several challenges, encompassing the significant disparities in terms of imaging resolutions between CT and CTCA, as well as slice misalignments due to patient motion between the exams.

1.2 Goals

Thus, the objective of this Dissertation is to integrate information from both imaging modalities to obtain comparable diagnostic information from non-contrast CT scans as is achieved with CTCA scans. This includes:

- Development of automatic frameworks for the alignment of CT and CTCA scans and implementing different super-resolution techniques.
- Validation of the developed methods using a dataset provided by the Centro Hospitalar Vila Nova de Gaia e Espinho (CHVNGE).

1.3 Document Structure

Besides the introduction, this document is structured in more five chapters:

Chapter 2 gives an overview of CAD, including crucial aspects of heart anatomy, the current diagnostic strategies for CAD, and the motivation of the thesis in this spectrum.

Chapter 3 provides a state of the art review of the current strategies of automatic methods for SR of non-contrast CT images. Additionally gives an overview of the main metrics to access the performance of the image reconstruction.

Afterwards, chapter 4 presents two distinct methods are presented as a preliminary work towards the SR reconstruction in cardiac CT imaging, interpolation and CNN-based methods with validation in a external private dataset from CHVNGE.

Chapter 4 focuses on advanced research to achieve super-resolved CT images at previously unavailable positions. It adapts the methods from the previous and tests autoencoders to enhance CT reconstruction integrating CTCA auxiliary information. A connected autoencoder approach for image separation is also presented.

Finally, the main conclusions, challenges, and future work are discussed in Chapter 6. This chapter presents the primary findings derived from this dissertation, along with suggestions for enhancing the developed algorithms and proposing future follow-up projects.

Chapter 2

Background

2.1 Heart Anatomy

The heart serves as the central conductor of the circulatory system, with its main role being as indispensable as life itself, driving blood throughout the entire network of blood vessels. It serves as a muscular pump, tirelessly and rhythmically contracting to distribute blood throughout the body, guaranteeing that every cell receives the oxygen and nutrients necessary for survival [9]. Any impairment in its function can lead to serious repercussions for overall health and quality of life. Therefore, a comprehensive understanding of the heart's anatomy, function, and the factors influencing its health is imperative for preserving cardiovascular well-being.

The internal cavity of the heart is divided into four chambers, two atria and two ventricles, with a wall of tissue called septum dividing the right and left side. The atria are thin-walled chambers that receive blood from the veins, while the ventricles are thick-walled chambers that pump blood out of the heart. These differences in thickness of the heart chamber walls are due to variations in the amount of myocardium present, which reflects the amount of force each chamber is required to generate.

In the circulatory system, the heart's four chambers work in unison to facilitate the flow of blood. The left atrium receives oxygenated blood from the lungs and sends it to the left ventricle, which propels this oxygen-rich blood to circulate throughout the body. The right atrium collects deoxygenated blood from the body and sends it to the right ventricle, which pumps the blood to the lungs for oxygenation [10]. The heart ensures the smooth progression of blood through a sequence of events that occur in a single heartbeat, represented in Figure 2.1, allowing for the exchange of gases in the lungs and the delivery of oxygenated blood to the entire body.

In its role as the unceasing conductor of the body's circulation, the heart necessitates a consistent flow of oxygen-rich blood to fuel its continuous contractions. This vital task falls upon the coronary arteries, which are responsible for providing the heart with the nourishment it needs to keep beating. The coronary arteries branch out across the heart's surface, forming a network that ensures every region of this remarkable organ receives the essential oxygen and nutrients required for its relentless pumping action [9]. Within the network of coronary arteries, the most prominent

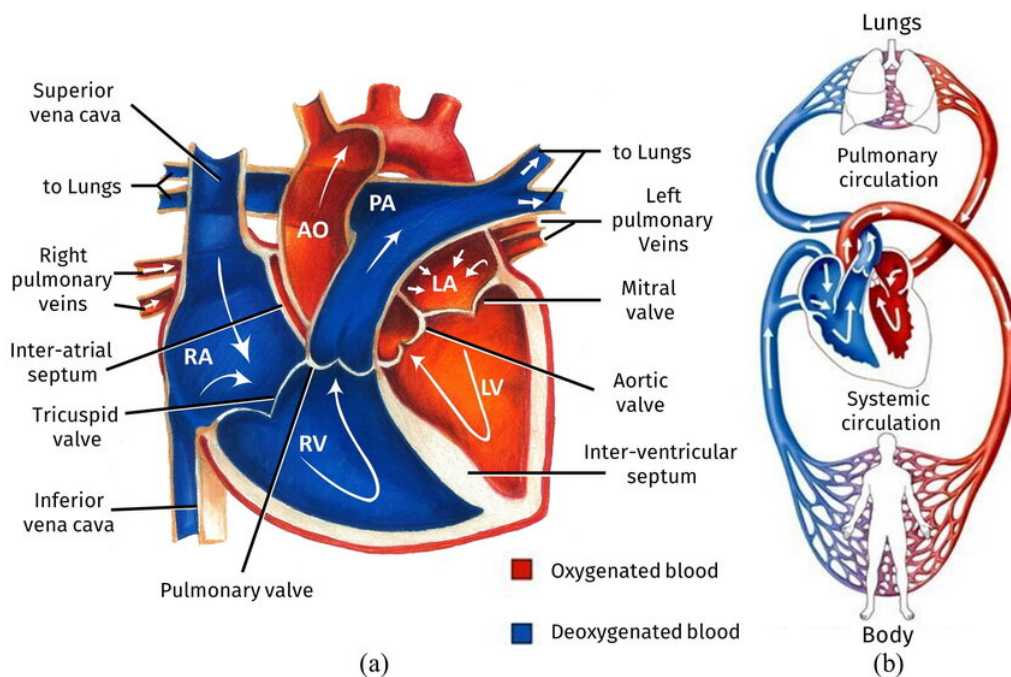


Figure 2.1: (a) Blood flow through the heart; (b) The heart inside the whole circulatory system. Image from [11]

are the left and right coronary arteries, which meander across the heart's surface, nurturing the life within. These arteries, along with their complex smaller branches, penetrate deep into the heart muscle, delivering the vital elements needed for its continuous, rhythmic contractions.

Coronary arteries are notable for their intricate constitution, essential for cardiac function. Their structure comprises three distinct layers in the arterial wall shown in Figure 2.2. The innermost layer, known as the intima, is formed by endothelial cells that line the artery's interior, allowing for smooth blood flow. The middle layer, called the media, consists of smooth muscle and plays a crucial role in regulating the artery's diameter, adjusting blood flow as needed. Lastly, the outer layer, the adventitia, is composed of connective tissue that provides structural support to the artery [12].

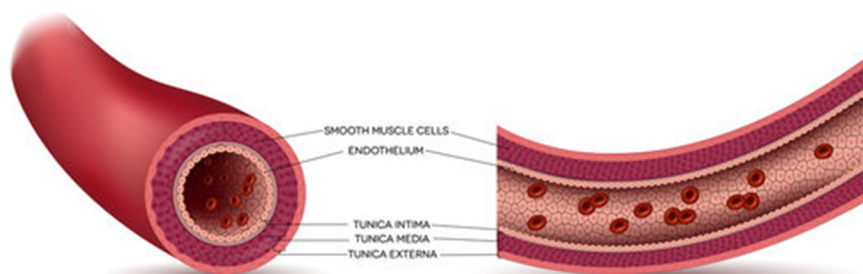


Figure 2.2: Cross Section of a Coronary Artery [13].

This complex constitution of the coronary arteries renders them robust enough to withstand

the oxygen and nutrient demands of the heart muscle, ensuring its uninterrupted function vital to the human body.

2.2 Coronary Artery Disease

2.2.1 Etiology

Coronary Artery Disease (CAD), also known as Ischemic Heart Disease (IHD), is a pervasive and globally prevalent cardiac ailment fundamentally associated with the progression of atherosclerotic plaque within the coronary arteries. This gradual process is driven by intricate mechanisms, and it leads to an inadequate supply of blood and oxygen to the myocardium. This insufficiency results from the occlusion of the coronary arteries create a critical mismatch between the demand and supply of oxygen [14].

Atherosclerosis, the defining feature of CAD, is a complex and persistent inflammatory condition characterized by the alteration of coronary artery structures. At its core, atherosclerotic plaque primarily comprises lipid deposits, predominantly sourced from low-density lipoproteins circulating in the bloodstream [15]. This accumulation is complemented by alterations in arterial smooth muscle cells and the generation of connective tissue elements.

CAD develops as a consequence of the gradual accrual of atherosclerotic plaque within the coronary arteries, as shown in Figure 2.3. This plaque formation is characterized by the deposition of calcium, fatty lipids, and abnormal inflammatory cells [16]. These deposits contribute to the stiffening of the artery walls and the subsequent narrowing of their inner lumen. Notably, the presence of calcium phosphate, particularly in the form of hydroxyapatite deposits within the muscular layer of blood vessels, plays a pivotal role in promoting arterial stiffness and initiating coronary arteriosclerosis [17].

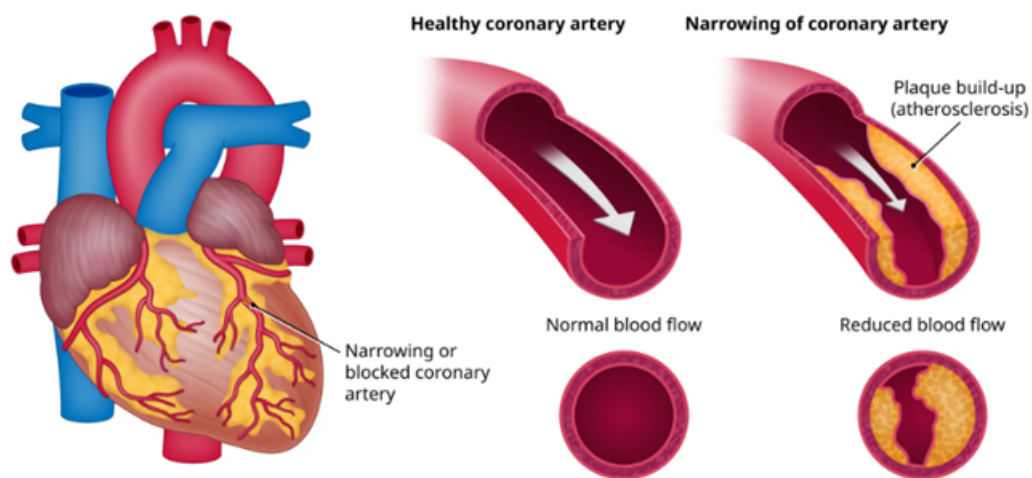


Figure 2.3: Illustration of the formation of the atherosclerotic plaque in CAD. Image from [18].

As these atherosclerotic plaques progressively accumulate, they obstruct the coronary arteries, limiting the flow of blood to the heart. This obstruction leads to a state of ischemia, wherein the heart's muscle cells are deprived of adequate oxygen. Ischemia is a precursor to myocardial infarction, more commonly known as a heart attack, wherein the heart muscle cells experience cellular death due to insufficient oxygen supply. The consequences of such an event include damage, cell demise, and the eventual formation of scar tissue within the heart muscle, with limited to no regeneration of heart muscle cells.

2.2.2 Epidemiology

CAD is a pervasive global health concern, standing as a leading cause of death worldwide. It results in over 164 million Disability Adjusted Life Years (DALYs) and 9 million annual fatalities globally, with an additional 7,100 lives lost each year in Portugal[19]. In the United States, the majority of cardiac deaths can be attributed to CAD arising from coronary atherosclerosis, being responsible for nearly 600,000 deaths annually [20].

CAD's influence is not confined to developed nations, it extends to the developing world, where it constitutes 2.2% of the overall global burden of disease and a remarkable 32.7% of all cardiovascular diseases [2].

Its impact on the global population is far-reaching, transcending geographic boundaries, gender, and age. In Europe, death rates stemming from both CAD and stroke tend to be higher in Central and Eastern Europe when compared to Northern, Southern, and Western Europe [21]. Furthermore, while it can affect individuals at any age, its prevalence significantly increases as individuals advance in age, with roughly a threefold increase in risk for each successive decade of life [22]. In 2020, it was estimated that approximately 20.1 million adults aged 20 and older were living with CAD.

CAD is influenced by a constellation of risk factors that contribute to its widespread prevalence. Lifestyle choices play a critical role, with factors such as smoking, obesity, poor dietary habits, chronic stress, physical inactivity, and excessive alcohol consumption significantly elevating the risk. Genetic predisposition also contributes, as individuals with a family history of CAD face heightened susceptibility [23]. Conditions like high blood pressure, diabetes, high cholesterol levels and depression further amplify the risk. Age is an undeniable factor, with CAD becoming notably more common as individuals advance in years, and males tend to exhibit a higher vulnerability than females.

2.2.3 Diagnosis

The diagnosis of CAD is a fundamental aspect of cardiovascular healthcare and medical imaging has become central in this endeavor. Traditionally, functional tests, including stress electrocardiography (ECG), exercise electrocardiogram (EKG), cardiac nuclear imaging with single-photon emission computed tomography (SPECT) or positron emission tomography (PET) and nuclear myocardial perfusion imaging (MPI) were used for this diagnosis [24], however they yield an

insufficient diagnostic accuracy for detecting obstructive CAD in terms of both sensitivity and specificity.

CT provides the basis for two coronary imaging techniques that convey very different information, capabilities that are not offered by any other non-invasive techniques. Thus, in recent years, the most common approach for CAD diagnosis consists of two anatomic tests, starting with Computed Tomography (CT) scan (or Non-Contrast CT- NCCT) and followed by a higher resolution Computed Tomography Coronary Angiography (CTCA) examination. This approach is essential in CAD diagnosis, each offering unique advantages [25].

Non-Contrast Computed Tomography (NCCT):

CT imaging, also known as a CT scan, is the first-line diagnostic test in the investigation of suspected CAD. This non-invasive modality provides a broad view of the coronary artery anatomy, facilitating the identification of calcifications that might obstruct blood flow. It serves as an initial screening tool to assess the general condition of the coronary arteries, being currently used for the analysis of Coronary Artery Calcium Score (CACS), which is considered to be the best marker of subclinical CAD [26]. In addition, it can provide important information with regard to epicardial adipose tissue (EAT), which is a marker of systemic inflammation and atherosclerotic cardiovascular disease risk.

However, it should be noted that while CT is valuable for initial assessment, it may not provide the level of detail necessary for a comprehensive CAD diagnosis.

Coronary Artery Calcium Score (CACS):

CACS represents a critical indicator of coronary atherosclerosis that can be used to measure calcified subclinical atherosclerotic burden, typically quantified using non-contrast CT scans [27]. This measurement provides insights into the presence and extent of calcium deposits within the coronary arteries (Figure 2.4), offering valuable information as a predictor of cardiovascular disease risk, including CAD.

The Agatston score is a commonly used method for CACS quantification, determined by the product of total calcified plaque area in millimeters squared and a quantized peak calcium density weighting factor, measured in HU. This score helps categorize individuals into various risk groups, aiding in risk stratification and informing clinical decisions. A higher CACS generally indicates a greater burden of coronary artery calcification, which is associated with an increased likelihood of experiencing CAD-related events [28].

However, while it effectively identifies the presence of calcification, it does not provide detailed information about the type or pattern of calcium deposition within the vessel walls. This limitation can impact its accuracy in predicting obstructive CAD and individual patient outcomes. Moreover, studies have shown that a high CACS may not necessarily correlate with plaque instability, and in some cases, plaque stabilization interventions can result in an increase in the CACS

[29].

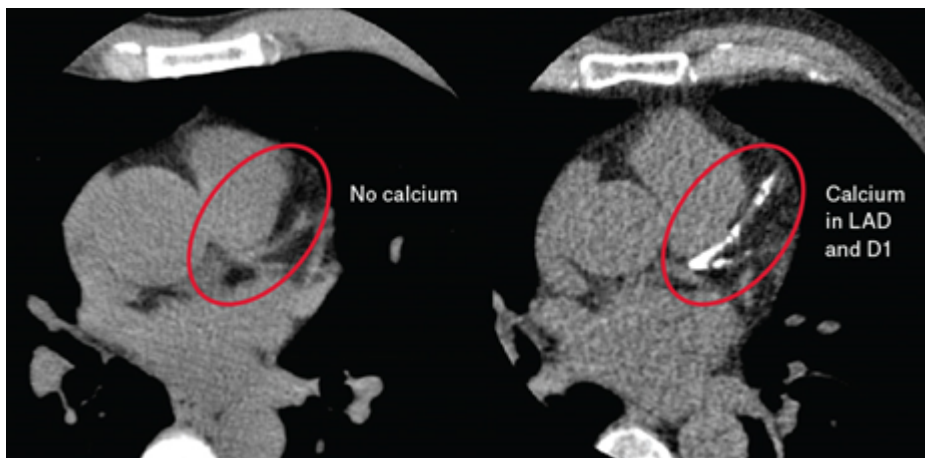


Figure 2.4: CACS exam without (left) and with (right) calcium in the arteries. Image from [27].

Computed Tomography Coronary Angiography (CTCA):

A unique feature of CTCA is its ability to provide a three-dimensional view of the arterial lumen, facilitating the detection and quantification of stenosis, the narrowed segments of coronary arteries. This technique requires the use of intravenous contrast and is performed after a non-contrast prospective ECG-triggered CT acquisition to plan the CTCA scan [30].

Stenosis quantification allows expert cardiac CT cardiologist or radiologist to classify patients based on the severity of arterial narrowing, following the CAD Reporting and Data System (CAD-RADS) guidelines [31]. The CAD-RADS is a standardized method for reporting and communicating coronary CT angiography findings and serves as a clinical decision support tool to guide subsequent patient management, assigning a numerical rating to test results what helps doctors assess the extent and severity of coronary artery blockages observed in the images. It categorizes findings into six levels, as shown in Figure 2.5.

CTCA can visualize arterial blockages, as stenosis in the coronary arteries, offering crucial insights about obstructions of the blood flow (Figure 2.6). This assessment is vital in gauging the severity of CAD and in pinpointing regions of the heart muscle that may be receiving insufficient blood supply due to coronary artery issues [32].

Overall, CTCA's capacity to capture dynamic aspects of coronary artery health provides clinicians with a comprehensive understanding of CAD pathology, aiding in accurate diagnosis and patient risk assessment. This detailed assessment is invaluable for guiding treatment decisions and interventions, ultimately leading to improved patient outcomes.

While CT and CTCA serve distinct purposes, their combined use offers a comprehensive CAD assessment. The initial CT scan provides a preliminary overview, identifying potential blockages and areas of concern. Following this, CTCA steps in to offer a more detailed and precise visualization of the coronary arteries, enabling the detection and quantification of stenoses. This combined

Table 4
CAD-RADS Reporting and Data System for patients presenting with stable chest pain.

Category	Degree of maximal coronary stenosis	Interpretation	Further Cardiac Investigation	Management considerations
CAD-RADS 0	0% (No plaque or stenosis)	Absence of CAD ^a	None	- Reassurance. Consider non-atherosclerotic causes of symptoms
CAD-RADS 1	1–24% (Minimal stenosis or plaque with no stenosis ^b)	Minimal non-obstructive CAD ^b	None	- Consider non-atherosclerotic causes of symptoms - P1: Consider risk factor modification and preventive pharmacotherapy - P2: Risk factor modification and preventive pharmacotherapy - P3 or P4: Aggressive risk factor modification and preventive pharmacotherapy
CAD-RADS 2	25–49% (Mild stenosis)	Mild non-obstructive CAD	None	- Consider non-atherosclerotic causes of symptoms - P1 or P2: Risk factor modification and preventive pharmacotherapy - P3 or P4: Aggressive risk factor modification and preventive pharmacotherapy
CAD-RADS 3	50–69% (Moderate stenosis)	Moderate stenosis	Consider functional assessment ^c	- P1, P2, P3 or P4: Aggressive risk factor modification and preventive pharmacotherapy - Other treatments (including anti-anginal therapy) should be considered per guideline directed care^d - When modifier 1+, consider ICA, especially if frequent symptoms persist after guideline-directed medical therapy
CAD-RADS 4	A - 70–99% stenosis or B - Left main $\geq 50\%$ or 3- vessel obstructive ($\geq 70\%$) disease	Severe stenosis	A: Consider ICA ^e or functional assessment B: ICA is recommended	- P1, P2, P3 or P4: Aggressive risk factor modification and preventive pharmacotherapy. - Other treatments (including anti-anginal therapy and options of revascularization) should be considered per guideline directed care^e
CAD-RADS 5	100% (total occlusion)	Total coronary occlusion or sub-total occlusion	Consider ICA, functional and/or viability assessment	- P1, P2, P3 or P4: Aggressive risk factor modification and preventive pharmacotherapy. - Other treatments (including anti-anginal therapy and options of revascularization) should be considered per guideline directed care^e
CAD-RADS N	Non-diagnostic study	Obstructive CAD cannot be excluded	Additional/alternative evaluation may be needed	

Figure 2.5: CAD-RADS Reporting and Data System for patients presenting with stable chest pain. Image from [31].

approach ensures that CAD diagnosis is thorough and accurate, providing essential information for clinicians to make informed decisions regarding treatment and intervention.

2.3 Joint CT/CTCA Analysis

In the UK's National Health Service (NHS), CTCA is recommended as the first-line investigation for patients with CAD suspicion, leading to a notable increase in its annual utilization[34]. However, this rise in CTCA referrals has sparked concerns regarding the increased exposure to ionizing radiation and iodine-containing contrast medium (ICCM), particularly among younger patients. CTCA procedures can subject patients to radiation doses exceeding 20 mSv, markedly higher than those from other types of chest CT scans [35]. The inherent risk of radiation exposure is widely acknowledged as harmful, with its effects accumulating over time, potentially leading to long-term risks such as cancer.

Recent data from CHVNGE revealed that among patients who underwent CTCA in 2019, 35% of patients showed no coronary plaque (CADS-RAD of 0) and 30% had only minimal disease (CADS-RAD of 1) indicating that more than 60% of patients were at low cardiovascular risk and likely unnecessarily exposure to ICCM and radiation [36]. This could potentially be avoided if similar diagnostic information could have been obtained from the non-contrast, low-dose radiation CT scans, since it offers a safer initial step in cardiac imaging with radiation doses below 3 mSv.

NCCT provides valuable information currently not used in clinical practice. Improving NCCT's diagnostic value through advanced image analysis tools could potentially rule out CAD in about 35% of patients and discharge them before undergoing CTCA, saving time, resources and avoiding unnecessary radiation exposure.

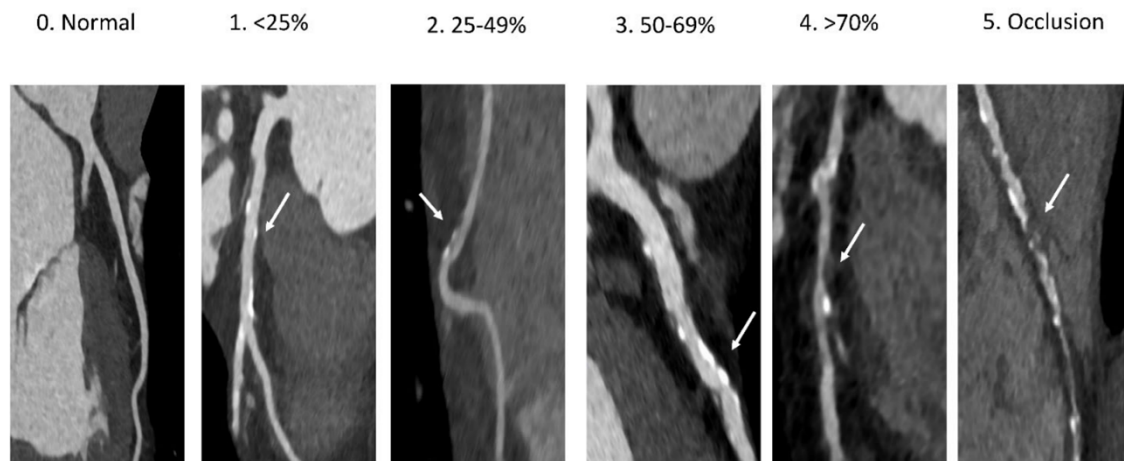


Figure 2.6: CTCA exams with the visualization of stenosis in the coronary arteries grading from 0–25% to 100% occlusions. Image from [33].

The integration of both CT and CTCA exams, facilitated by deep learning methods, offers a promising advantage by providing a comprehensive and synergistic approach to cardiovascular disease assessment. This integration allows for a holistic evaluation of coronary health. By leveraging these complementary modalities, it becomes possible to achieve a more accurate and nuanced understanding of CAD, ultimately leading to improved diagnostic precision, enhanced treatment planning, and the potential for highly personalized patient care. However, this strategy is constrained by several technical and clinical issues.

One of the primary limitations lies in the differences in image resolutions between CT and CTCA exams. CTCA, being an angiographic examination, typically has higher axial resolution (typically in the range of 0.5 mm slice thickness [37]), allowing for detailed visualization of the coronary arteries. Conversely, non-contrast CT scans may have lower resolutions (typically in the range of 2.5 to 3 mm slice thickness [38]), which can affect the accuracy of detecting lesions or obstructions in the coronary arteries. This discrepancy in image quality can hinder the effective fusion of information from both exams. Furthermore, patient motion between CT and CTCA exams is an additional challenge since small shifts or respiratory movements can cause misalignments between image slices from the two exams, further impairing the ability to combine and interpret the data accurately and seamlessly.

For the last-mentioned limitation, the misalignment between the exams, Spectral CT technology emerges as a promising solution to address this issue in the future. Spectral CT, also known as dual-energy CT, is an advanced imaging technique that uses two separate X-ray photon energy spectra to acquire images [39]. This allows the acquisition of multiple distinct attenuation datasets from the same scan, what can be used to acquire a variety of image types, including material decomposition, which is particularly useful in CTCA scans where iodinated contrast is employed.

By allowing the acquisition of CT and CTCA images within a single examination, Spectral CT holds the potential to reduce alignment errors [40]. However, it is essential to note this technology, despite its potential benefits in medical imaging, is currently limited by certain drawbacks. The

primary obstacles are its limited accessibility and high cost. As a relatively recent and continuously evolving medical technique, the initial setup and investment required for these systems can pose significant barriers for numerous healthcare facilities, consequently limiting patient access [41]. Moreover, this technology may not be readily available in all geographic regions, exacerbating disparities in patient healthcare. Nevertheless, while it may not be widely available at the moment, it stands as an exciting and prospective solution in the near future.

Thereby, given the limitations presented as well as the lack of existing solutions to address both, the motivation for developing super-resolution and alignment frameworks between CT and CTCA exams is imperative. Super-resolution (SR) aims to enhance the quality of CT images by increasing their resolution to match that of CTCA, thereby enabling more effective and precise integration of information from both imaging modalities to obtain comparable diagnostic information from non-contrast CT scans as is achieved with CTCA scans.

These innovative approaches not only have the potential to improve CAD detection and diagnosis, enabling more targeted screening strategies that can appropriately identify patients who may not require a CTCA, thereby reducing unnecessary exposure to ICCM, minimizing scanning time and costs, and optimizing healthcare resources. They also pave the way for a range of other clinical applications, such as CAD classification based on multiple imaging modalities and high-quality image generation to assist clinicians and researchers in their assessments. Therefore, the development of CT and CTCA SR and alignment frameworks is an urgent necessity in the pursuit of advancements in the field of cardiovascular diagnosis.

Chapter 3

Automatic Super Resolution of CT Images

In general, SR can be divided into three categories: interpolation-based SR [42, 43], reconstruction-based SR [44, 45] and learning-based SR. Super resolution methods have come a long way from simple interpolation techniques to sophisticated deep learning models. Among these algorithms, the learning-based methods have gradually become a research interest in this field for producing results that are superior compared to other algorithms.

Traditionally, the interpolation methods have been dominant on the SR research, such as bicubic interpolation and Lanczos resampling, among others. These simple and fast methods involve the straightforward pixel interpolation to create higher-resolution images, where the missing pixel values are estimated by, for example, the weighted average of the nearby pixels with known values [46].

Reconstruction-based SR methods emerged with the advancement of techniques such as CT and other medical imaging modalities. These methods employ reconstruction algorithms, including iterative reconstruction or frequency domain transformations like Fourier or wavelet transform, to enhance resolution and facilitate the acquisition of sharper and more detailed images in medical and scientific applications [47].

In recent years, learning-based methods have revolutionized SR research by harnessing the power of deep neural networks and machine learning techniques. These approaches have the ability to generate high-resolution images by learning from large datasets, effectively mapping low-resolution inputs to high-resolution outputs [48]. The introduction of deep learning-based SR methods marked a significant breakthrough, enabling the modeling of complex patterns and the production of higher-quality images.

Researchers have explored a variety of methods in the field of SR, ranging from the early adoption of Convolutional Neural Networks (CNN) to more recent advancements like Generative Adversarial Networks (GAN). It is important to note that during the research underlying this literature review, two distinct types of techniques for achieving image super-resolution (SR) using deep learning (DL) have been identified: non-axial SR and axial SR [49].

- Non-Axial, or In-Slice, SR methods primarily focus on enhancing the resolution in the plane of the image, which means that its main objective is to improve the ability to distinguish details and features within the two-dimensional space of the image. Typically, these methods operate on the existing image slices obtained from CT scans and work to augment the clarity and level of detail within each individual image slice [50]. This is the most common type of SR, with a vast number of studies addressing this topic in various ways. Nevertheless, in line with the research objectives, the primary focus in this state-of-the-art review will be confined to the topic of axial super-resolution methods.
- Axial SR methods concentrate on enhancing the resolution along the axial direction, which corresponds to the direction along the length of the patient's body. This approach typically entails reconstructing or enhancing information along the z-axis (depth or longitudinal direction) to improve the ability to distinguish details along the patient's body or throughout the entire scan volume. Axial SR methods typically focus on improving the resolution in the direction that the scanner acquires images, such as enhancing the distance between slices in the z-direction [51]. Achieving axial SR may involve employing interpolation algorithms or advanced image reconstruction techniques, allowing for the creation of thinner axial images than those originally acquired and thereby improving the resolution along the z-axis.

This chapter provides an overview of different methods and techniques used for fully automatic super-resolution in CT/CTCA. Within the first section, an exploration of the techniques and methodologies recently employed in the axial SR of medical images is undertaken, with the main focus being CT images. This section is organized into two sub-sections specially dedicated to the presentation of deep learning approaches for axial super-resolution, namely CNNs-based techniques, and GANs-based techniques. Additionally, a dedicated section provides an exposition of the evaluation metrics typically employed to assess the quality of the SR achieved through these techniques. In the final section, an analysis of the chapter's conclusions and the limitations inherent in the discussed approaches are presented, thereby providing a thorough perspective on the prevailing landscape in this research area.

3.1 State of Art

The research methodology for conducting the literature review aimed at understanding the state of the art in this field of study began with a focused approach using specific keywords. Initially, the search targeted "Axial or Z-axis Super Resolution" or "Axial or Z-axis Interpolation/Reconstruction", specifically within cardiac CT imaging. However, due to limited results in this specific area, the scope was expanded to encompass a broader range within CT imaging, not solely limited to cardiac applications. Subsequently, the search was further broadened to include various other medical imaging modalities. This gradual expansion allowed for a comprehensive exploration of studies related to super resolution and axial interpolation techniques across different CT regions and other

medical imaging modalities.

In the current landscape of research in this field, the majority of existing works can be broadly categorized into two distinct DL techniques: CNNs and GANs. These studies have been summarized chronologically in Table 3.1 to provide a succinct summary of their integration into this review, highlighting their respective corresponding findings and the best outcomes.

Within the document, these studies are structured in ascending order of complexity, facilitating a detailed analysis of the progression and evolution of research methodologies and outcomes in the targeted field.

Table 3.1: Summary of Deep Learning algorithms for super resolution.

Authors	Number of Data	Image Modality	DL's method	PSNR	SSIM	Others metrics
Yukai Wang et al. [52]	5 sets	Rock samples CT	CNNs	40.00	0.987	—
Viren Jain [53]	29 scans	Microscopy, CT and MRI	CNNs	—	—	Visual Assessment.
Jun-Young Park et al. [54]	65 scans	Brain CT	CNNs	41.68	—	NRMSE
Cheng Peng et al. [55]	120 scans	T1 MR brain	CNNs	32.22	0.9441	DSC
Akira Kudo [56]	407 scans	CT	GANs	35.73	0.933	Visual Assessment.
Cheng Peng et al. [57]	853 scans	CT	CNNs	40.82 (liver dataset)	0.9821 (liver dataset)	—
Jiawei Li et al. [58]	1400 slices	Torso cross-section CT	GANs	19.27 (MSE loss)	0.9117 (MSE loss)	—
Zhuoqun Yuan et al. [51]	1,000 scans	OCT	GANs	28.58	0.6853	MSSSIM, MOS and Visual Assessment
Heng-Sheng Chao et al. [59]	35 sets	Chest CT	CNNs	39.021	0.9227	—

3.1.1 CNNs

CNNs, a class of deep, feed-forward artificial neural networks, have emerged as a dominant force in computer vision due to their ability to automatically learn and extract intricate features from images. Their hierarchical architecture, which includes convolutional layers for feature extraction and pooling layers for spatial downsampling, enables CNNs to effectively capture complex patterns in visual data. This has sparked significant interest and research efforts directed towards applying CNNs to tasks involving super-resolution[60].

This study stands as one of the early research initiatives in the field, dedicated to achieving axial super-resolution in CT images, however, in a distinct context, specifically in the field of geological CT images. In any case, the article discusses the potential applicability and relevance of this technique to medical CT images. In Wang’s study [52], a novel network is proposed, the Three-Dimensional Super Resolution CNN (3DSRCNN), for achieving voxel super resolution in CT images through deep learning, in the x, y, and z directions. The network architecture consists of 12 layers of 3D-CNN (Figure 3.1), with the initial layers extracting low-frequency information and the subsequent ones learning the mapping relationship between low-resolution and high-resolution volumetric blocks. Yukai Wang et al. employ strategies such as adjustable learning rates, residual learning, gradient clipping, and stochastic gradient descent with momentum (SGD) to optimize training and mitigate convergence issues. The 3DSRCNN was trained, with Mean Squared Error (MSE) loss function, on 5 sets of CT images with varying upscale factors and resolutions, allowing it to perform multi-scale super resolution with a single model. Evaluation metrics include PSNR and SSIM, demonstrating superior performance compared to conventional methods.

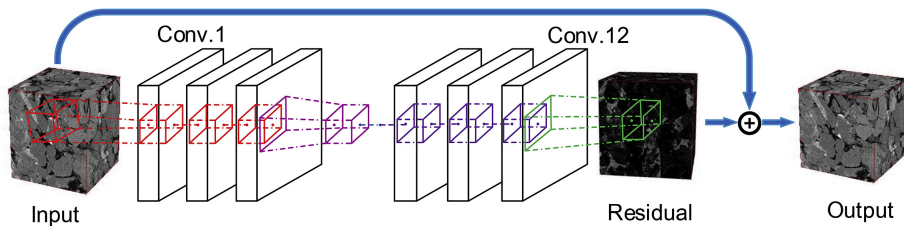


Figure 3.1: Network structure of 3DSRCNN contains 12 layers 3D-CNN [52].

Heng-Sheng Chao et al. [59] presents a novel approach for generating high-resolution CT images using deep learning-based SR techniques with the main objective to enhance the resolution of thin-slice CT images by fusing two sets of thick CT images (5 mm of thickness in both axial and coronal series) into a single interpolated 3D LR image and then applying 4 different SR models to the third orthogonal plane: SRCNN (Super-Resolution Convolutional Neural Network), VDSR (Very Deep Super Resolution), SRResNet (Generator Network for SRGAN) and EDSR (Enhanced Deep Residual Network). Image preprocessing involved 3D image registration and averaging of axial and coronal thick slices to generate LR DICOM data for model training, represented with the diagram in Figure 3.2, and testing. The study highlighted the superior performance of the

EDSR model over other SR algorithms, as evidenced by quantitative measurements such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM).

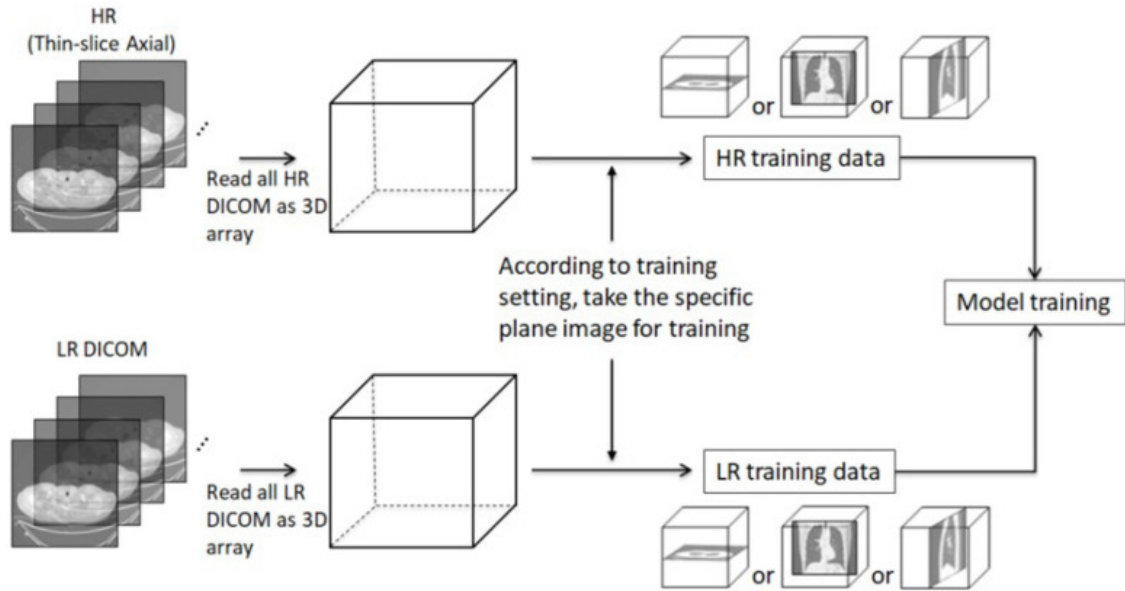


Figure 3.2: Diagram of model training for a specific plane [59].

In Viren Jain’s approach [53], the focus is on image interpolation and alignment, utilizing feedforward neural networks trained with adversarial loss functions. The primary contributions include a fully convolutional network architecture designed to interpolate image sections from adjacent sections and producing better-aligned versions of volume data. Specifically, the research addresses the interpolation of individual 2D image sections within 3D image volumes (Figure 3.3), emphasizing the synthesis of missing or damaged image data for a particular section based on data from neighboring sections. The network architecture is deliberately kept simple but effective, trained with an adversarial loss function that involves a generator network and a discriminator network, where the generator produces interpolated slices while the discriminator distinguishes between real and generated slices. This approach seeks to achieve pixel-wise alignment and interpolation by balancing the strength of both the discriminator and generator. The method is implemented using the ADAM optimizer and TensorFlow on a GPU and the evaluation process combines quantitative metrics and human to assess image alignment quality, with the primary goal being the generation of smooth and coherent boundaries in resliced volume data. In sum, this approach showcases the potential of adversarial neural networks in addressing the challenges of image interpolation, alignment, and super-resolution, highlighting their utility in various medical imaging applications.

Junyoung Park’s work [54] also presents a CNN approach to enhance the resolution and contrast of CT images acquired with thick slices, addressing the common trade-off between resolution

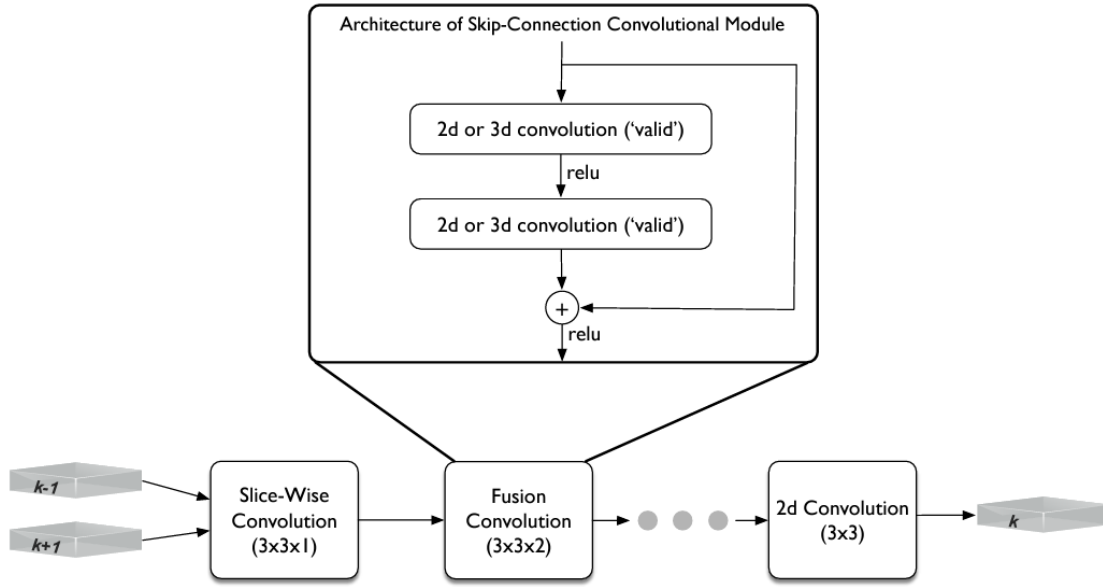


Figure 3.3: Architecture of the slice-interpolation network [53].

and noise in CT imaging. The study applies a modified U-Net architecture to learn a sophisticated end-to-end mapping between low-resolution CT images (thick slices) and their high-resolution counterparts (thin slices). This deep CNN architecture used in this study consists of multiple layers of convolutional and pooling operations, followed by fully connected layers (Figure 3.4). The network is trained on a large dataset of low- and high-resolution brain CT images, using a loss function that minimizes the difference between the predicted and ground-truth high-resolution images. Specifically, 52 CT studies serve as the training set, and 13 CT studies are allocated for testing, ensuring robustness through a five-fold cross-validation approach. The evaluation of the method employs image quality metrics such as PSNR and normalized root mean square error (NRMSE), and it is compared to other state-of-the-art super-resolution methods. The results show that the proposed method outperforms other methods in terms of both quantitative metrics and visual quality. Overall, this study demonstrates the potential of deep CNN for improving the resolution of CT images and enhancing the accuracy of medical diagnoses.

Cheng Peng et al. [55] introduces an approach for slice interpolation in brain magnetic resonance (MR) scans, called marginal super-resolution (MSR), aimed to interpolate between sparsely sampled slices in anisotropic MR volumes in order to improve the resolution in the axial direction. Leveraging 2D CNNs, the proposed approach consists of two stages: marginal super-resolution (MSR) and two-view fusion and refinement (TFR), represented in Figure 3.5. In the MSR stage, high-quality estimates of the interpolated slices are obtained by extrapolating contextual information from sagittal and coronal axes. In the TFR stage, the sagittally and coronally super-resolved volumes are further refined using information from the axial axis. The fusion network combines the features from the super-resolved volumes, and the refinement network improves the quality

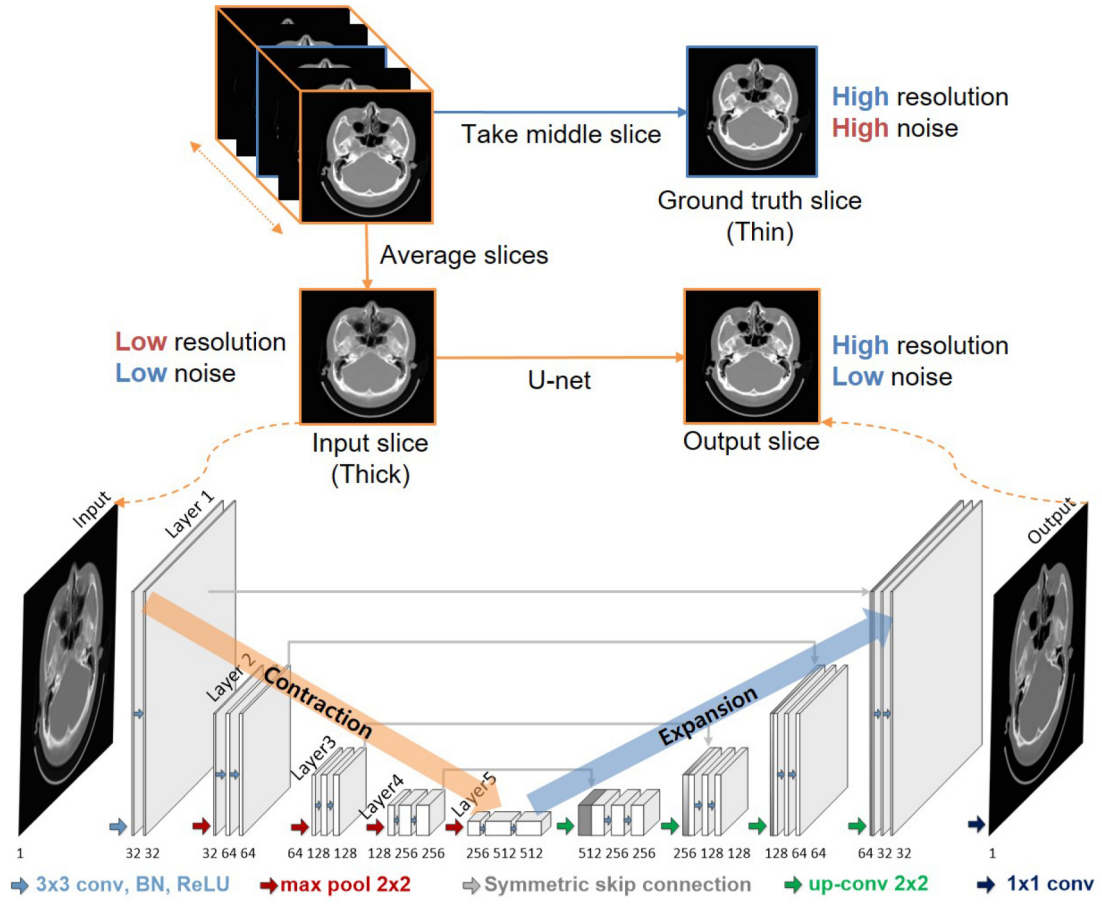


Figure 3.4: CNN architecture with data flow [54].

of the interpolation by learning the structural variations along the axial direction. This study is conducted on a dataset comprising 120 T1 MR brain scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI), with evaluations based on PSNR, SSIM, and gray/white-matter segmentation. The average value obtained for PSNR was 41.68, while the average value for SSIM was 0.9441, and, regarding the segmentation results evaluated with the Dice Similarity Coefficient (DSC) metric, these were 0.9021 for the gray matter segmentation and 0.8593 for the white matter. Its results demonstrate the superior performance of the proposed method over traditional interpolation techniques and baseline CNN models, showcasing its potential for enhancing brain volume estimation from sparsely sampled MR scans.

While this study was conducted using brain magnetic resonance images, it serves as a valuable foundation for the author's subsequent research endeavor aimed at adapting the model for use with CT images, which aligns with our specific area of interest. This development is noteworthy because it underscores the potential to adapt models that perform well in other medical imaging modalities to effectively apply them to CT images. This adaptability holds significant advantages, given the relatively limited research in this particular area concerning CT images.

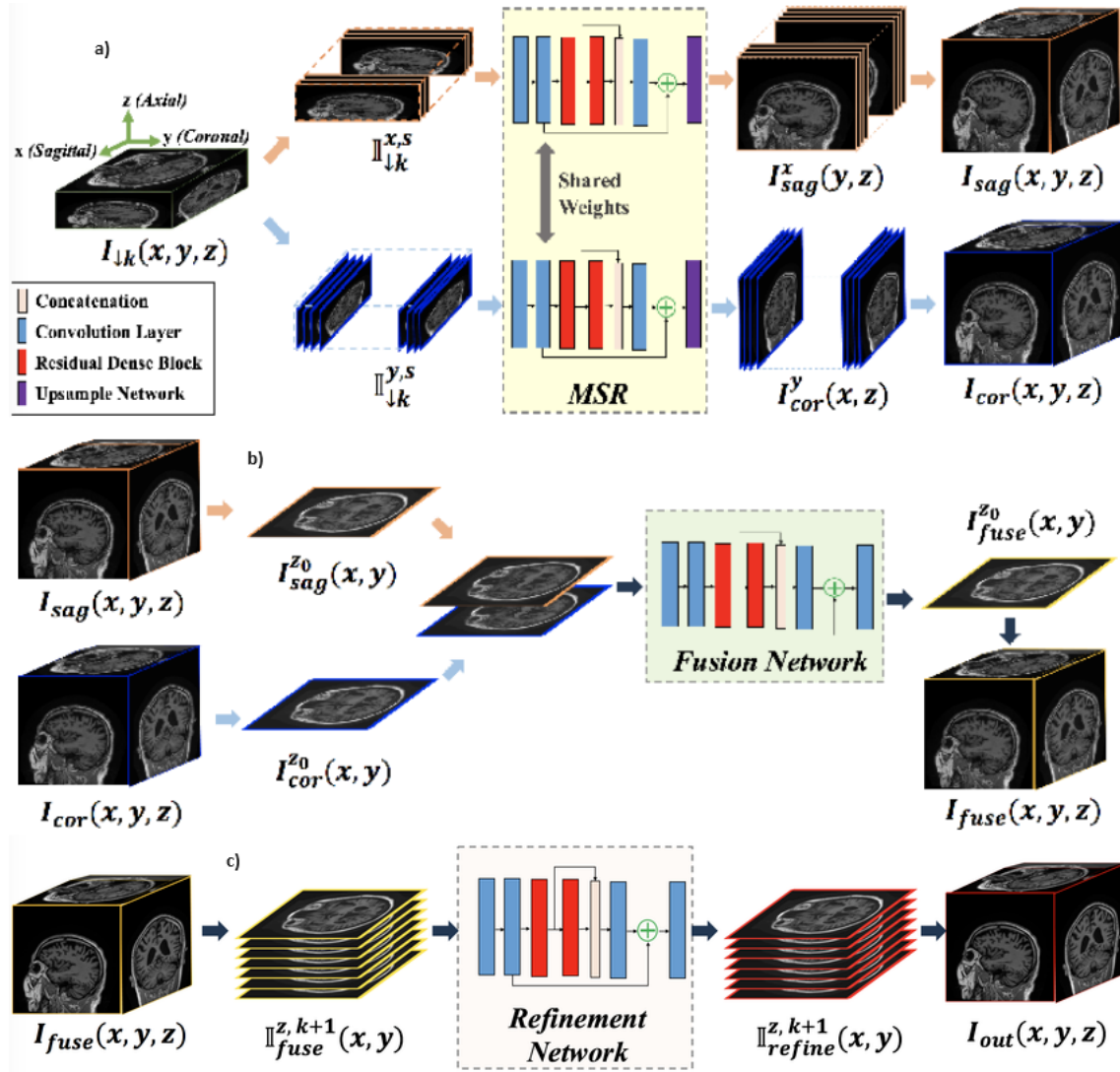


Figure 3.5: a) Marginal Super-Resolution Pipeline. b) Two-view Fusion Pipeline. c) Refinement Pipeline [55].

In another study, Cheng Peng et al. [57] introduce a Spatially Aware Interpolation NeTwork (SAINT) approach for medical slice synthesis to alleviate the memory constraint that volumetric data poses. SAINT utilizes voxel spacing information to provide desirable levels of details, and adaptively determines the upsampling factor, overcoming the limitations of other super-resolution methods. The network consists of two main stages (Figure 3.6): the first being the Anisotropic Meta Interpolation (AMI) mechanism, which leverages voxel spacing information to provide desired levels of detail and dynamically determine the resolution upsampling factor. Based on the previous mentioned work, the second stage is the Residual-Fusion Network (RFN), which addresses inconsistencies between slices in the axial direction resulting from applying AMI to 3D CT images and incorporates information from the third axis for improved 3D context modeling. This work demonstrates the effectiveness of SAINT through extensive evaluation on 853 CT scans

from four public datasets that contain liver, colon, hepatic vessels, and kidneys featuring a mix of slice thicknesses ranging from 1mm to 4-5 mm. The results showcase its superior performance quantitatively on independent datasets and unseen upsampling factor, by using performance metrics such as PSNR and SSIM.

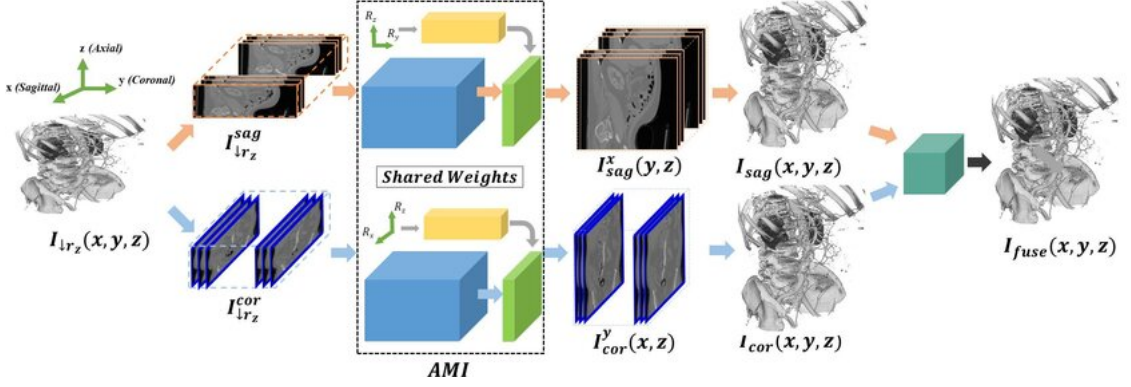


Figure 3.6: The overall pipeline of Spatially Aware Interpolation NeTwork (SAINT). For visualization purpose, the volumes are rendered in 3D based on their bone structures [57].

3.1.2 GANs

Finally, automated synthesis of image data has recently been advanced by the invention of generative adversarial networks (GANs) [61]. GANs are advanced neural networks comprising a generator and a discriminator that compete during training to generate realistic data, such as images. They excel in tasks like image synthesis and style transfer by creating high-quality outputs indistinguishable from real data.

In Akira Kudo’s research [56], a cutting-edge approach is introduced, leveraging conditional GANs (cGANs) to enhance the resolution of CT images, addressing the limitations of thick-slice CT images by transforming them into virtual thin-slice images. The proposed method employs 3D fully convolutional neural networks for generating high-resolution images from various fields of view (Figure 3.7), demonstrating superiority in image quality through metrics such as PSNR, SSIM and Visual Turing Test (VTT). In this technique, the generator and discriminator models are based on CNNs with 4x4x4-sized kernels, employing batch normalization and LeakyReLU activation. The generator follows a U-Net-inspired encoder-decoder architecture with trilinear interpolation for upsampling and the discriminator takes thick and thin images (1 channel each) and slice information (8 channels) as input and includes optional self-attention. A Degradar procedure simulates thin-to-thick image degradation through Gaussian smoothing, spline interpolation, and random Gaussian noise. The discriminator is conditioned on a vector (w) with information about input data type, slice interval, and slice thickness, totaling 8 conditional channels. The training

dataset had 354 CT data (head:99, chest:98, abdomen:100, legs:57), focusing on 1 mm slice intervals from 8 mm. Testing included 53 CT data (head:12, chest:16, abdomen:15, leg:10), confirming successful 1 mm slice interval image generation across diverse slice spacings and field-of-view settings.

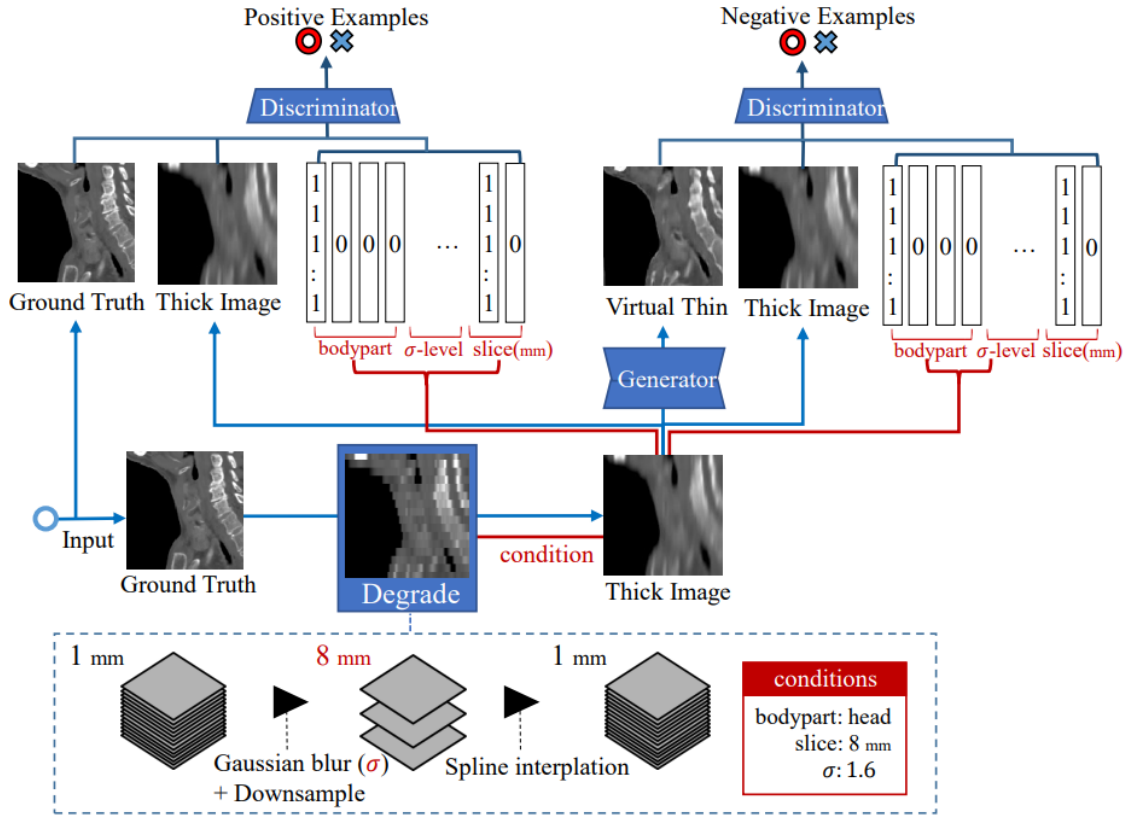


Figure 3.7: Adversarial training framework of thick-thin slice translation on CT images. In each training iteration, the thin slice input data is randomly degraded to simulate thick slice data. The generator input is a 1 mm spline interpolated 3D thick slice image itself and the discriminator input is the generated Virtual Thin image with slice condition including body part information with degraded parameter scales [56].

Jiawei Li et al. [58] present a novel approach to medical image interpolation called the High-Resolution Interpolation Network (HRINet). The main objective is to address the problem of image interpolation in the medical field, where the distance between consecutive slices in biomedical volume images is often significantly larger than the in-plane pixel size. HRINet combines the idea of Adversarial Cross-Modal Autoencoder (ACAI) with Generative Adversarial Networks (GANs) and proposes an alternative supervision method by applying supervised and unsupervised training alternatively to raise the accuracy of human organ structures in CT while keeping high quality. The network uses multidimensional interpolation between feature maps and disentangles the latent representation to generate high-resolution images. HRINet's architecture consists of a generator network and a discriminator network, Figure 3.8, where the generator uses a se-

ries of convolutional layers to upsample the low-resolution input image, while the discriminator evaluates the realism of the generated image. To train and evaluate HRINet, Jiawei Li et al. harnessed a dataset comprising 1400 CT slices derived from 10 distinct patients, extracted from the torso cross-section, including abdomen and pelvis regions. The training of HRINet involves two distinct loss optimization techniques: MSE loss and a perceptual-based loss. The evaluation incorporates metrics such as PSNR, SSIM, RMSE and compares the performance of different models. Experimental results presented in the study demonstrate a substantial quantitative and qualitative improvement, therefore the proposed HRINet has the potential to enhance the quality of 3D reconstruction and visualization in the medical field.

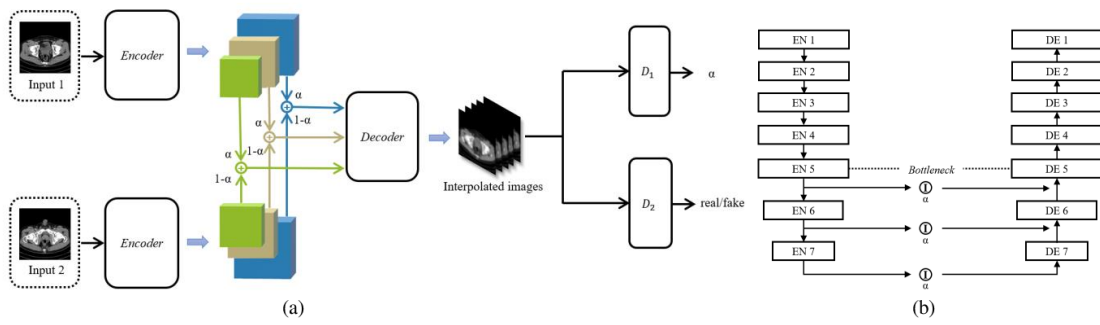


Figure 3.8: The overview of our model. An encoder is trained to produce low-level feature maps, which are interpolated with interpolation coefficient α and the decoder tries to disentangle the latent representation and map it into data space. D_1 is the discriminator trying to predict α , and D_2 is trained to distinguish whether the input is real or not. (b) Network architecture of the autoencoder part. "EN" or "DE" stands for encoder or decoder block [58].

During this research, an additional study was encountered in which axial super-resolution was achieved, albeit in a different medical imaging modality, Optical Coherence Tomography (OCT). In the work conducted by Zhuoqun Yuan et al. [51], the primary objective is to enhance the axial resolution of Optical Coherence Tomography (OCT) images, which are crucial in medical fields such as ophthalmology. The proposed approach introduces residual-in-residual dense block network (RRDBNet), Figure 3.9, to super-resolve LR OCT images acquired from their custom spectral domain OCT (SD-OCT) system. Using a dataset of 1,000 matched axial LR and HR OCT image pairs obtained from their custom SD-OCT system, the network is implemented in PyTorch, trained with the Adam optimizer, with a combination of loss functions, including adversarial loss and perceptual loss and evaluated using metrics like PSNR, SSIM, multi-scale-structural similarity index measure (MSSSIM), and the subjective evaluation mean opinion score (MOS). The results demonstrate that deep learning methods, particularly the RRDBNet trained with various loss functions, yield the best super-resolution for OCT images. The experiments also suggest that these deep learning techniques exhibit good generalization and robustness across different OCT systems.

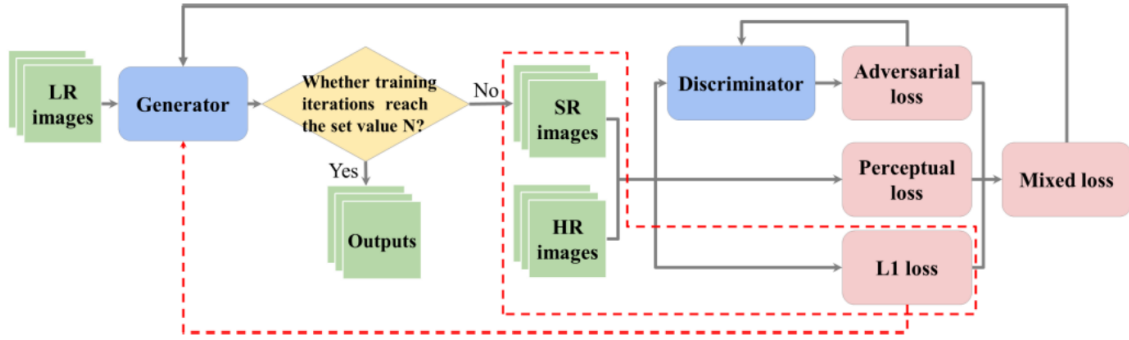


Figure 3.9: Layout of the deep learning network [51].

3.2 Evaluation Metrics

The selection of evaluation metrics for SR image reconstruction carries significant weight as it profoundly influences our capacity to assess the model's performance and efficacy. Various evaluation metrics are designed to capture unique facets of image quality, and the choice must be tailored to the precise objectives. In this section, three main categories of common evaluation metrics used in medical image reconstruction will be addressed, such as: paired metrics, unpaired metrics, and perceptual metrics.

3.2.1 Paired Metrics

Paired metrics focus on assessing the quality of image reconstructions by comparing them directly to a known ground truth image. This comparison allows for a quantitative analysis of how closely the reconstructed image matches the true or expected image, facilitating precise and objective evaluations of the reconstruction accuracy.

- **Mean Squared Error (MSE)** MSE is a widely used metric in medical image quality assessment. It calculates the average of the squares of the differences between pixel values in a reconstructed image and the original image. A lower MSE indicates greater similarity between the images, which is vital in medical applications where image fidelity matters.

$$MSE = \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N} \quad (3.1)$$

where N is the total number of pixels, y_i represents the pixel value in the original image, and $\hat{y}_i$ represents the pixel value in the predicted image [62]. MSE is a widely used metric in medical image quality assessment. It calculates the average of the squares of the differences between pixel values in a reconstructed image and the original image. A lower MSE indicates greater similarity between the images, which is vital in medical applications where image fidelity matters.

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (3.2)$$

where N is the total number of pixels, y_i represents the pixel value in the original image, and $\hat{y}_i$ represents the pixel value in the predicted image [62].

- **Normalized Root Mean Square Error (NRMSE)** It is essentially a modification of the MSE that considers the varying scales of pixel values within images. In medical imaging, pixel intensities can differ significantly due to factors like contrast adjustments or different imaging modalities. Consequently, NRMSE addresses this issue by normalizing the MSE with respect to the observed range of pixel values in the image, allowing for the comparison of image quality while accounting for differences in pixel value scales.

$$NRMSE = \frac{\sqrt{MSE}}{Max - Min} \quad (3.3)$$

where Max is the maximum pixel value in the image and Min the minimum pixel value in the image [63].

- **Peak Signal-to-Noise Ratio (PSNR)** PSNR stands as a crucial metric in the realm of medical image quality assessment. Derived from the MSE, PSNR reveals the relationship between the maximum pixel intensity and the power of distortion present within an image. This logarithmic representation is particularly valuable because medical images possess an extensive dynamic range, spanning from their largest to smallest potential values, which can vary based on their quality.

In practice, the PSNR emerges as the go-to tool for assessing the quality of reconstructed images generated by lossy image compression codecs. It delineates the original data as the signal and the compression or distortion-induced errors as noise. The PSNR serves as an approximate measure, mirroring human perception of reconstruction quality relative to the performance of compression codecs.

$$PSNR = 10 \cdot \log_{10} \left(\frac{R^2}{MSE} \right) \quad (3.4)$$

where R is the maximum possible pixel value [64].

It quantifies the quality of an image by comparing the maximum possible signal intensity to the power of the distortion (MSE) in a logarithmic scale (base 10), providing a measure of how well the image has been preserved or reconstructed. In general, a higher PSNR indicates that the reconstruction is of higher quality.

- **Structural Similarity Index (SSIM)** The SSIM method is a perception-based model used to assess the structural similarity between a reconstructed image and its original counterpart. It takes into account factors such as luminance, contrast, and image structure. This metric

quantifies structural patterns of pixel intensities, particularly among neighboring pixels, after adjusting for luminance and contrast variations. Because the human visual system is highly adept at perceiving structural details, the SSIM quality metric closely aligns with subjective quality assessments. The SSIM score ranges from -1 to 1, with 1 indicating maximum similarity.

Originally developed to predict perceived quality in digital images and videos, SSIM has proven valuable in the medical field. It offers a precise measurement of the resemblance between original and reconstructed medical images, thus improving diagnostic accuracy and facilitating medical research.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3.5)$$

where μ_x is the pixel sample mean of x , μ_y is the pixel sample mean of y , σ_x^2 is the variance of x , σ_y^2 is the variance of y , σ_{xy} is the covariance of x and y , $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ the dynamic range of the pixel-two variables to stabilize the division with weak denominator, L is the dynamic range of the pixel-values, and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

The SSIM formula is based on three comparison measurements between the samples of x and y : luminance (l), contrast (c) and structure (s). The individual comparison functions are:

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (3.6)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (3.7)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \quad (3.8)$$

with, in addition $c_3 = c_2/2$ [65].

- **Multi Scale Structural Similarity Index (MSSSIM)** The MSSSIM is an advanced extension of the SSIM that offers a more comprehensive evaluation of medical image quality by operating on multiple scales, considering both local and global image features. It assesses structural similarities between a reconstructed medical image and its original form by considering luminance, contrast, and image structure, just like SSIM. However, MSSSIM goes further by decomposing the image into multiple scales and computing SSIM values at each scale. These individual scale SSIM values are then combined to provide a more detailed and nuanced assessment of image quality. MSSSIM is particularly useful in medical imaging,

where subtle details and local variations can be crucial for diagnosis and research. It allows for a more accurate evaluation of image fidelity and aids in making informed decisions regarding the suitability of medical images for clinical use or research applications [66].

The MSSSIM is defined in several steps:

1. Divide the original image (X) and the reconstructed image (Y) into multiple scales (e.g., using Gaussian pyramids).
2. For each scale, calculate the SSIM between images X and Y, obtaining multiple SSIM values (SSIM1, SSIM2, etc.).
3. Combine these SSIM values for each scale into a single weighted MSSSIM value, considering the relative importance of different scales.
4. The general formula for MSSSIM is:

$$MSSSIM = (SSIM1^\alpha) * (SSIM2^\beta) * (SSIM3^\gamma) * \dots \quad (3.9)$$

Where SSIM1, SSIM2, SSIM3, etc., are the SSIM values calculated for each scale and α, β, γ , etc., are weights that determine the relative importance of each scale.

- **Dice Similarity Coefficient (DSC)** The DSC is a commonly used metric in the field of medical image analysis and computer vision. It measures the similarity between two sets of data, such as binary masks or segmentations of an image, where higher values indicate greater overlap and thus higher agreement between the segmented regions [67].

The formula for the DSC is given by:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (3.10)$$

where A and B are the two sets of data being compared, and |A| and |B| represent the number of elements in each set.

The DSC ranges from 0 to 1, where 0 indicates no overlap between the regions and 1 indicates complete overlap. This metric is particularly useful in segmentation tasks where it is important to quantify the accuracy and agreement of segmented regions with respect to a reference region, such as in medical diagnostic images.

3.2.2 Unpaired Metrics

Unpaired metrics are designed to assess the quality and realism of artificially generated images without the need for a direct reference to real images. These metrics provide a quantitative framework to evaluate the faithfulness and perceptual quality of the reconstructed images, enabling a rigorous assessment of their accuracy and realism in various medical imaging applications.

- **Inception Score (IS)** The IS is an objective metric used to assess the quality and diversity of generated images, specifically synthetic images output by GANs. It calculates two components: the average classification probabilities assigned to the generated images by a pre-trained neural network, often InceptionNet, to measure image quality, and the entropy of class distributions, which evaluates the diversity of generated images.

The IS formula calculates the average Kullback-Leibler divergence between the conditional distribution ($p(y|x)$) and the marginal distribution ($p(y)$) for all generated images. A higher value indicates greater diversity in the classes assigned to the images, which is indicative of greater diversity in the generated images. This metric provides valuable insights into the performance of generative models, assisting in the determination of their effectiveness in producing realistic and varied images.

$$IS(G) = \exp(\mathbb{E}_x[KL(p(y|x)||p(y))]) \quad (3.11)$$

where E_x represents the expected value (average) over all generated images and $KL(p(y|x)||p(y))$ denotes the Kullback-Leibler divergence between two distributions: (1) $p(y|x)$, the conditional distribution of class probabilities (y) given the generated images (x), in other words, the probability of a class given a specific image; (2) $p(y)$, the marginal distribution of class probabilities (y) in the set of generated images, in other words, it is the probability of a class in the set of images [68].

However, interpreting the values obtained from IS is not always straightforward, and this metric has several limitations. Firstly, the IS algorithm is designed to work on small, square image sizes. Secondly, IS measures image diversity, meaning that a limited sample size, such as only a single type of image or the same image produced multiple times, can result in an artificially high Inception Score due to insufficient variation to judge true diversity. Additionally, the IS is calculated against a pre-trained dataset within the image classification network that encompasses around 1000 image classes, so if the generated image does not fit within these predefined classes, the IS will yield an artificially low score because there is no corresponding pre-trained data to compare against.

- **Fréchet Inception Distance (FID)**

FID is a metric designed for evaluating the quality of generated images, primarily developed to evaluate the performance of GANs. Introduced in 2017, it has become the standard measure for evaluating generative models since around 2020. Unlike its predecessor, IS, FID incorporates real images, ground truth, in its evaluation alongside computer-generated images. This dual assessment aims to more accurately reflect human perception, making FID a preferred metric for gauging the fidelity and diversity of generated images.

The FID score measures the distance between feature vectors computed for real and generated images, the *Fréchet* distance, which is a statistical measure of the difference between two probability distributions. Thus, it provides a quantification of the similarity between

the two groups in terms of statistical properties related to computer vision features extracted from raw images.

$$\text{FID}(G, R) = \|\mu_R - \mu_G\|^2 + \text{Tr}(\Sigma_G + \Sigma_R - 2(\Sigma_G \Sigma_R)^{0.5}) \quad (3.12)$$

where μ_R and μ_G refer to the feature-wise mean of the real reference (R) and generated images (G), Σ_R and Σ_G refer the covariance matrix of the features extracted from the real reference images and generated images [69].

In contrast to the IS, which primarily evaluates the distribution of generated images, FID takes a step further by comparing the distribution of generated images to the distribution of a set of real images from the target domain.

Lower FID scores indicate a higher similarity between the two image groups, with a perfect score of 0.0 signifying identical distributions. FID is a valuable tool in the evaluation of GAN-generated images, with lower scores often correlating with higher image quality.

3.2.3 Visual Assessment

Visual Assessment is a crucial aspect of medical image quality evaluation, involving a qualitative analysis performed by radiologists or experts. This assessment relies on the experienced eyes and judgment of human observers who evaluate various aspects of reconstructed images, such as clarity, fidelity, and utility. Unlike quantitative metrics, visual assessment provides insights into the perceived quality of medical images, capturing nuances that may not be quantifiable by mathematical algorithms alone. Radiologists and experts play a pivotal role in determining the suitability of images for diagnostic purposes, ensuring that they convey the necessary information accurately and effectively. This metric serves as a complementary approach to quantitative metrics, allowing for a holistic evaluation of image quality and its impact on clinical decision-making and medical research [70].

Visual assessment is particularly valuable for interpreting the quantitative results obtained, aiding in the identification of issues more readily by combining the interpretation of metric outcomes with visual analysis of the results.

3.3 Summary

In summary, this chapter has undertaken a comprehensive exploration of automatic methods for SR of non-contrast CT images. It begins by examining traditional SR approaches, including interpolation and reconstruction methods, which have shown some failures in restoring satisfactory images due to their lack of adaptability. Interpolation-based techniques, such as linear, bilinear, and bicubic methods, tend to reconstruct images with blurred edges and sparse textures, making them not sufficient for restoring missing details. On the other hand, reconstruction algorithms,

while capable of producing higher quality images than interpolation-based methods, require multiple images and precise alignment, posing practical challenges [71].

In recent years, the spotlight in the field of SR methods has shifted towards learning-based methods. Specifically, deep learning-based SR methods has showcased remarkable performance in various computer vision tasks, driven by convolutional neural networks (CNNs) and generative adversarial networks (GANs), have emerged as a breakthrough, enabling the modeling of complex patterns and the generation of higher-quality CT images.

In the context of axial super-resolution, convolutional neural networks CNNs and GANs offer distinct yet complementary approaches. CNNs excel in extracting features from images and demonstrate greater stability during training, making them suitable for tasks requiring detailed processing of existing data; they are supported by extensive literature in our specific problem domain. However, CNNs have more limitations in generating new details or enhancing the realism of images. On the other hand, GANs are renowned for their capability to produce high-quality and remarkably realistic images, making them ideal for super-resolution tasks by synthesizing details and textures not present in the original data. However, they encounter challenges such as intensive computational complexity and training instability since training GANs involves optimizing two models—the generator and discriminator—simultaneously in a dynamic setting. Ensuring the discriminator neither dominates too quickly nor becomes too weak is crucial for effective learning and to prevent issues like mode collapse, requiring meticulous adjustments throughout the training process [72].

Initially, challenges were encountered in locating studies within this specific realm of super-resolution. While there is an extensive body of literature related to non-axial super-resolution, the domain of axial super-resolution remains relatively underexplored. It is important to note that certain limitations and gaps were identified within the existing literature. Specifically, no methods specifically applied to cardiac CT images could be found, which represents a valuable area for this research. Furthermore, the reliance on paired metrics in the reviewed studies, with datasets having ground truth images for SR reconstruction, highlights the need for exploring unpaired metrics to evaluate models without ground truth data.

In the final section of this research, an analysis was conducted on three main categories of common evaluation metrics used in medical image reconstruction: paired metrics, unpaired metrics, and perceptual metrics. This assessment aimed to identify the most suitable metrics for evaluating this work.

Overall, moving forward in this research, the goal is to address these existing gaps and contribute to the advancement of axial super-resolution techniques in cardiac CT imaging. This will be approached through two primary strategies, detailed in the subsequent subsections: the application of interpolation-based methods and, mainly, CNN-based methods. The focus on learning-based techniques, particularly CNNs, is driven by their adaptability, frequent use in the field and their proven success in the context of medical imaging, as highlighted in this literature review.

Chapter 4

Intermediate CT Reconstruction

This chapter introduces two distinct methods as a preliminary work towards SR reconstruction in cardiac CT imaging. In this initial stage, it was considered addressing the SR problem using only CT data. Specifically, the planned approach involves utilizing input CT data in an interleaved manner, skipping every other slice, resulting in a reduced-resolution dataset. To elaborate, the input CT data are processed such that, in each iteration, two consecutive slices are provided while the intermediate slice is omitted. The goal is for a computational method to receive these two slices as input and generate the missing intermediate slice as output. This process aims to enhance the axial resolution of the CT data by reconstructing slices that are not present in the input data. Moreover, since the complete original CT data is available, the method's effectiveness can be evaluated by directly comparing the reconstructed slice with its corresponding ground truth. Hence, this approach not only aims to improve the axial resolution of the input CT data but also enables direct evaluation by comparing the reconstructed slices against the original slices in the full-resolution CT dataset, serving as an important initial step. To provide a clearer understanding, the following schematic diagram (Figure 4.1) illustrates this process.

The chapter begins with an overview of the dataset used for the validation of the developed methods, highlighting its main characteristics. The methods explored and tested in this chapter include an interpolation-based approach and a U-Net CNN for learning-based SR, which are well-established in SR tasks as discussed in existing literature. Both methods aim to achieve the same objective: the reconstruction of the intermediate slice from the two provided input slices. These methods are evaluated to assess their effectiveness in improving axial resolution in cardiac CT imaging, forming the foundational elements of our initial approach to SR. By adopting this streamlined approach, this chapter aims to provide a comprehensive overview of the methodologies employed, their respective implementations, and the findings obtained through their application in the field of cardiac CT imaging.

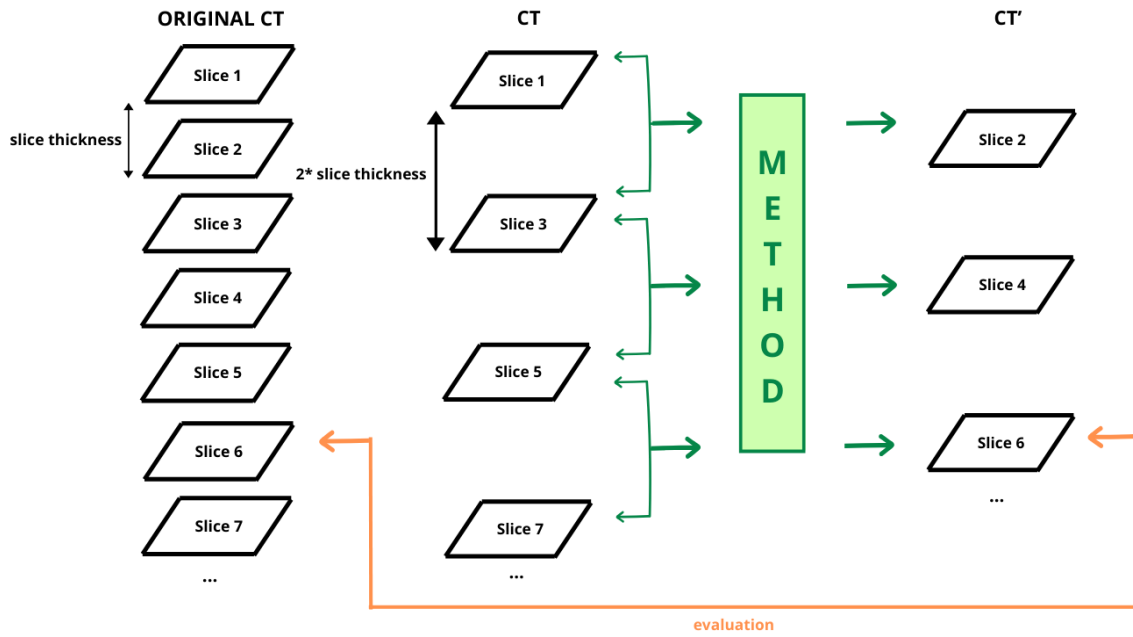


Figure 4.1: Schematic diagram illustrating the process of SR reconstruction using interleaved CT data.

4.1 Data

The CHVNGE dataset is a subset comprising data from 203 patients who underwent both CT and CTCA examinations for the purpose of CAD screening. This dataset was collected at the CHVNGE facility located in Vila Nova de Gaia, Portugal. It is essential to note that all patient-specific information in the dataset was anonymized before being subjected to analysis for the research study.

The dataset encompasses CT and CTCA scans for each patient, which are provided in the DICOM format and possess a dimension of 512x512 pixels. An example of a CT slice, as well as its corresponding CTCA slice, for the same patient from the CHVNGE dataset, is shown in Figures 4.2 and 4.3. All CT and CTCA scans were acquired using a Siemens Somatom Force scanner. The CT scans, in particular, were performed with a slice thickness of 1.5 mm, while the CTCA scans exhibit varying slice thicknesses ranging from 0.4 mm to 1 mm, offering finer resolution essential for detailed coronary artery imaging.

For the methods implemented in this chapter, only the CT data are employed as input. This data was randomly arranged for training, validation, and testing by creating 5 folds. Each fold included input data pairs, i.e. the two consecutive slices, and their respective targets, their original intermediate CT slice, organized to accommodate a specific number of patients, specifically 40 per fold. This strategy ensured the initiation of a new fold after processing each set of 40 patients. Establishing these folds was essential for an organized system, aiding easy access to training data and facilitating a streamlined 5 fold cross-validation process. The 5-fold cross-validation is a

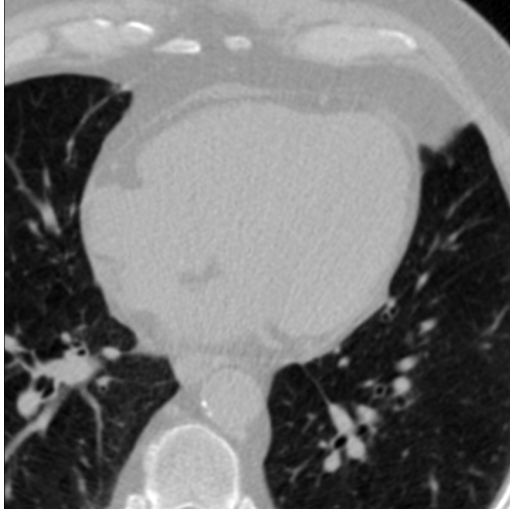


Figure 4.2: Slice of CT exam from the CHVNGE dataset.

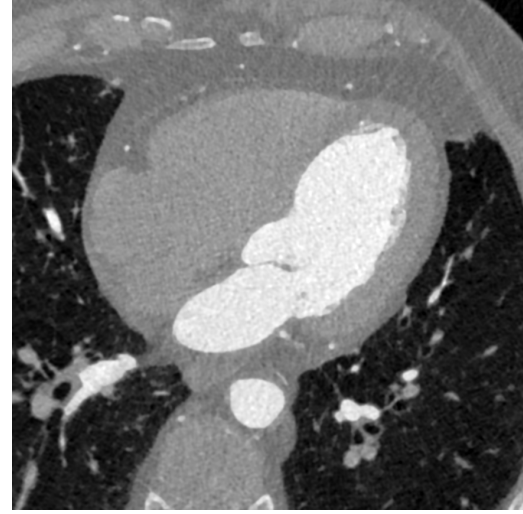


Figure 4.3: Corresponding slice of CTCA exam.

widely employed technique for model evaluation and validation [73]. It involves partitioning the dataset into 5 equally sized folds, with one fold serving as the validation set while the remaining 4 folds constitute the training set. This process is iterated 5 times, each time using a different fold as the validation set and the remaining folds for training. By averaging the performance metrics obtained from each iteration, a more reliable estimate of the model's performance can be obtained.

4.2 Methodology

4.2.1 Interpolation-Based CT Reconstruction

In the initial phase, the main goal was to develop a straightforward method for CT image SR, which will serve as a baseline for evaluating more advanced methods. As previously mentioned in Chapter 3, the simplest techniques originally employed for SR tasks often rely on interpolation-based methods [74].

An interpolation algorithm based on the direct averaging of the two input slices was developed initially. This algorithm aims to reconstruct an intermediate slice at the position positioned between two input slices, representing the lower and upper bounds around the target reconstructed slice in the output. Specifically, the algorithm reconstructs CT slices at positions where the corresponding real CT slices are available.

The scheme represented by Figure 4.4 provides a simple and illustrative example of how this specific method operates. Here, a pair of CT slices is given as the input, and the algorithm incorporates these slices to compute the interpolation using averaging to reconstruct the intermediate slice as the output.

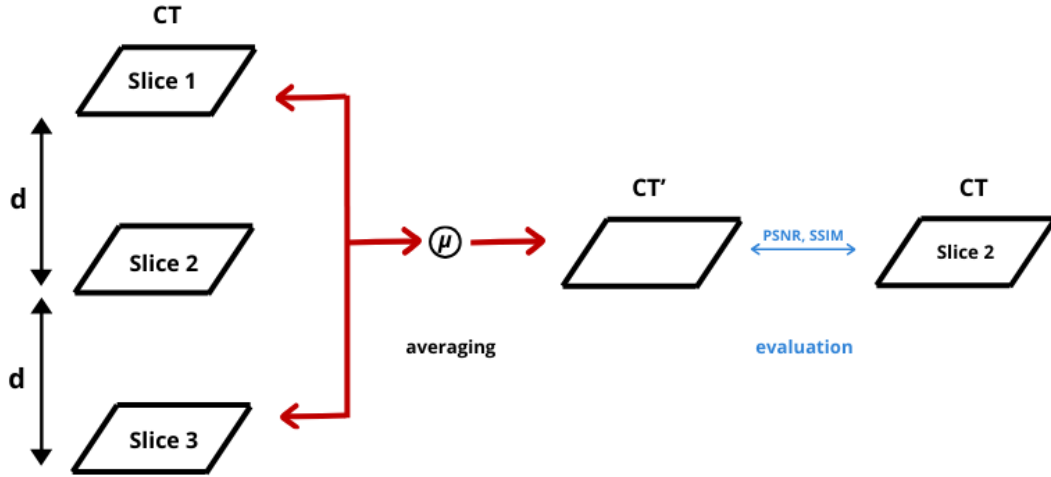


Figure 4.4: Scheme of the Interpolation-Based CT Reconstruction method.

4.2.2 CNN-Based CT Reconstruction

Comparatively, it was opted to test a learning-based technique involving CNNs. The main goal remained aligned with the initial interpolation objective, constructing a network capable of taking two input slices and training it to autonomously generate the intermediate slice, as represented in the diagram within the Figure 4.5.

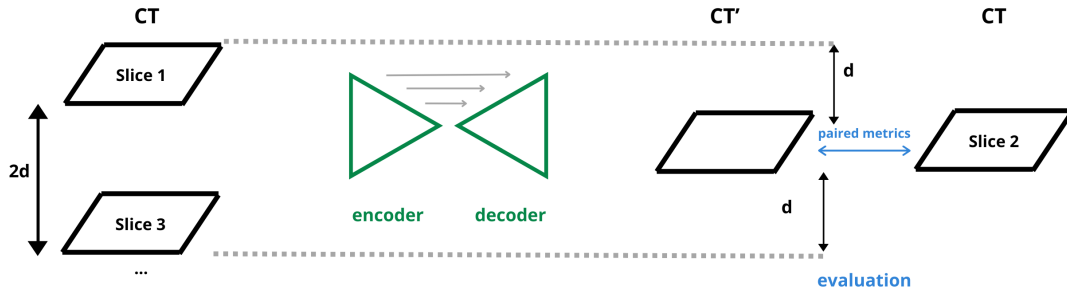


Figure 4.5: General outline of the CNN-Based CT Reconstruction model training goal.

Model Architecture

For the implementation of the model, a U-Net architecture was used [75]. The U-Net is characterized by its symmetrical U-shaped configuration, composed of a downward path, the encoder, and an upward path, the decoder, interconnected by skip connections between corresponding levels of both paths. The central idea of the U-Net is to enable the capture of contextual and spatial information at different scales by combining the top-down and bottom-up paths.

In this specific context, the U-Net is used to reconstruct an image at the output from a pair of input images. This implies two input channels and one output channel. The activation function utilized across all layers, except in the final one, is ReLU, along with the application of batch normalization and typically employing convolution operations. It's worth noting that the architecture

is two-dimensional (2D) and employs the 'transposed' method for upscaling. Figure 4.6 schematically represents the architecture of the U-Net utilized in this image reconstruction context.

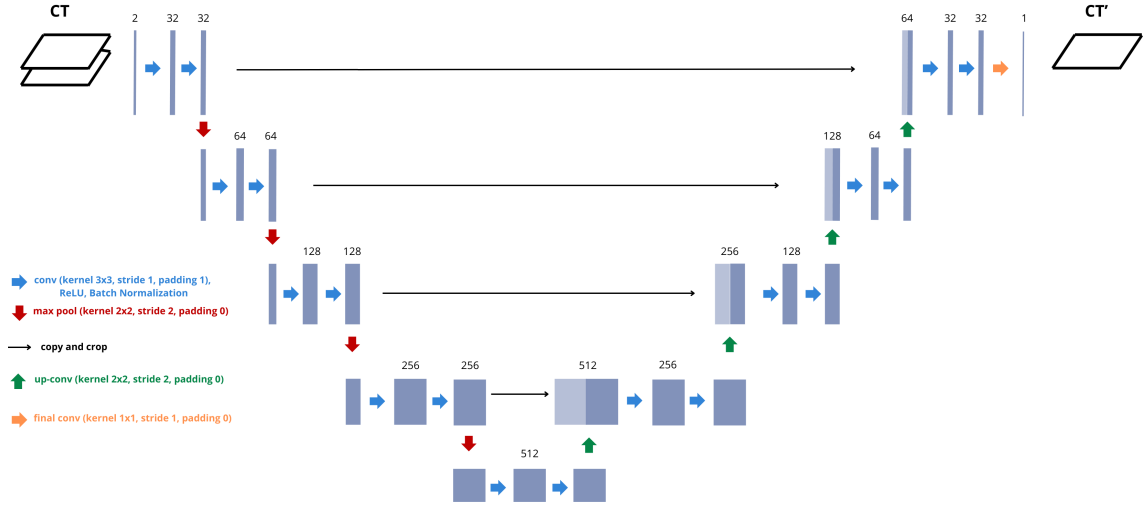


Figure 4.6: U-Net architecture used.

• Encoder

The downward path, also known as the encoder, comprises 5 down blocks. Each block consists of two convolutional layers with a kernel of 3×3 , a stride of 1, and padding of 1. Subsequently, a ReLU activation function is applied, followed by a Batch Normalization layer. Finally, a max-pooling operation is performed with a kernel size of 2, a stride of 2, and padding of 0 to reduce the spatial resolution.

Within each down block, the spatial resolution of the input is decreased while progressively increasing the number of channels (depth) of the image, determined by the number of filters, starting with an initial quantity of 32 filters. The encoding blocks are responsible for reducing the data dimensionality and capturing high-level features of the image.

• Decoder

The upward path, known as the decoder, consists of 4 up blocks, where each block comprises an upsampling layer, or transposed convolution with a kernel size of 2 and a stride of 2, followed by 2 convolutional layers. These convolutional layers have a kernel of 3×3 , a stride of 1, and padding of 1, utilizing the 'same' convolution mode. Each convolutional layer is succeeded by a ReLU activation function and a Batch Normalization layer.

Every up block receives inputs from both the corresponding level in the descending part, encoder, and the corresponding level in the ascending part, decoder. Within each ascending block, a concatenation operation is performed on the feature maps obtained in the ascending phase with the corresponding feature maps from the descending path. This process aims to preserve crucial spatial details.

The goal is to merge contextual information from the descending levels with high-resolution details from the ascending levels, enabling precise reconstruction of the image. The decoding blocks are specifically responsible for increasing the spatial resolution of the data and reconstructing the original image from the learned high-level representations.

- **Final Layer**

At the culmination of the ascending path, the final stage of the model incorporates a concluding 1x1 convolutional layer to map the feature maps to the desired number of output channels, which in this instance is 1, utilizing a kernel size of 1, a stride of 1, and a padding of 1.

Specifically in this model, a hyperbolic tangent (Tanh) activation function [76], multiplied by 1000, is employed, to produce the final output and ensure that the final output values range between -1000 and 1000.

Data Setup

The initial data preparation phase involved extracting three consecutive slices from each CT scan and segregating them into two distinct groups: one comprising input slice pairs and the other containing the corresponding real intermediate slices (target).

Data Preprocessing and Augmentation

Before training the network, all data underwent pre-processing steps. Initially, the HU values of the images were limited according to the range of [-1000, 1000] HU.

Data augmentation techniques were also employed in the training and validation sets in order to increase the variability of the data to enhance the model's ability to generalize and perform well on new and unseen data [77]. These techniques consisted of image resizing, random cropping, rotations within a range of 10 degrees and horizontal or vertical flips.

Training Details

Throughout the training process, a batch size of 6 samples per batch was used. The Adam optimizer was employed to adjust the model's weights, with a learning rate of 1e-4. An early stopping callback method was also used to stop training when the validation loss did not improve for 20 consecutive epochs and the maximum number of epochs without improvements was set as 300. The deep learning model was trained using PyTorch and an NVIDIA GTX 1080 GPU.

The model was trained using three different loss functions: MSE Loss, SSIM Loss, and a combined loss that incorporates both MSE and SSIM losses. Each of these loss functions was employed in separate training sessions to evaluate their impact on model performance.

The MSE Loss computes the mean of squared differences between generated image values and real image values, aiming to minimize differences between predicted pixels and the target.

$$MSELoss(y_i, \hat{y}_i) = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \quad (4.1)$$

where y_i represents the intensity of a pixel in the target image (the real image of the intermediate slice), $\hat{y}_i$ represents the intensity of a pixel in the output image produced by the model (the reconstructed intermediate slice) and N denotes the total number of pixels in the image.

Conversely, the SSIM Loss measures the similarity between two images, the $\hat{y}$ (the output image) and y (the target image), considering luminance, contrast, and structure, with the objective to minimize dissimilarity.

$$SSIMLoss(y, \hat{y}) = 1 - SSIM(y, \hat{y}) \quad (4.2)$$

where,

$$SSIM(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + c_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)} \quad (4.3)$$

With μ_y and $\mu_{\hat{y}}$ being the mean intensities of y and $\hat{y}$, respectively, σ_y^2 and $\sigma_{\hat{y}}^2$ the variances of y and $\hat{y}$, respectively, $\sigma_{y\hat{y}}$ the covariance between y and $\hat{y}$, and c_1 and c_2 are constants, $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, where L , the dynamic range of pixel values in the image, is equal to 2000 since the images were preprocessed to the range from -1000 to 1000.

The Combined Loss is defined with a factor to balance the contributions of MSE and SSIM losses to leverage the benefits of both pixel-wise accuracy and structural similarity, aiming to stabilize the training process and achieve optimal performance.

$$CombinedLoss(y, \hat{y}) = factor \cdot MSE(y, \hat{y}) + SSIM Loss(y, \hat{y}) \quad (4.4)$$

The factor used in our case was $\frac{0.06}{1100}$, where 0.06 represents the average value across folds around which the validation loss stabilized during training with SSIM Loss. Similarly, 1100 represents the average value around which the validation loss stabilized during training with MSE Loss. This factor ensures that both losses contribute equally within the same range of values, thereby balancing their influence during training in the Combined loss function.

4.3 Evaluation

For the evaluation process, since the real CT slices corresponding to the positions of the reconstructed slices are available, it was feasible to assess the quality and fidelity of the method's results

by comparing them to existing references using paired metrics. Specifically, Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) were selected as evaluation metrics.

SSIM, as mentioned in 3.2.1 assesses the structural similarity between two images by considering human perception of similarity based on luminance, contrast, and structure, where values closer to 1 indicate a higher structural similarity between the compared images. On the other hand, PSNR quantifies the relationship between the maximum signal intensity and the mean squared error between images, where a higher PSNR value indicate lower distortion between the compared images.

These metrics were computed for reconstructed images to assess their fidelity compared to real intermediate slices across all CT scans for each patient. This analysis provides a comprehensive understanding of how well the methods replicates real data.

Additionally, the Fréchet Inception Distance (FID) was also calculated. FID evaluates the realism of the reconstruction methods by comparing the distributions of features extracted from the set of generated images to those from the set of real intermediate slices, the ground truth. Unlike SSIM and PSNR, FID does not require a reference image for each synthetic image, making it a valuable metric for assessing the overall reconstruction quality between the generated and real image sets.

In summary, the objective of computing these metrics is to thoroughly evaluate and validate the quality, fidelity, and realism of the methods across the entire dataset of CT scans. This assessment provides quantitative insights into how well the generated images replicate the characteristics of the real intermediate slices, supporting the method’s effectiveness in clinical applications.

4.4 Results and Discussion

This section presents the results of intermediate CT reconstruction using the two developed methods: interpolation-based and CNN-based. Table 4.1 summarizes the quantitative results obtained for both methods, where bold indicates the best result for each metric.

Table 4.1: Comparison of Interpolation and U-Net with different loss functions results.

Method	SSIM	PSNR (dB)	FID
Interpolation	0.944 ± 0.0304	35.744 ± 3.08	8.018
U-Net (MSE Loss)	0.994 ± 0.0064	42.787 ± 4.52	6.532
U-Net (SSIM Loss)	0.985 ± 0.0134	39.547 ± 4.15	10.828
U-Net (Combined Loss)	0.995 ± 0.0065	42.268 ± 4.40	10.166

All the results obtained were promising and provided significant insights into the reconstruction of intermediate CT slices using both interpolation-based and U-Net-based methods.

As shown in Table 4.1, the interpolation-based method achieve an overall average SSIM of 0.944 ± 0.030 and a PSNR of it was 35.74 ± 3.08 . Additionally, the FID score obtained was measured at 8.0182.

Based on the obtained SSIM and PSNR values, it appears that the interpolation method employed to reconstruct the intermediate slice has yielded high-quality results. In general, the SSIM values fall between 0.76 to 0.98, indicating a high degree of structural similarity between the interpolated slices and the ground truth. Regarding PSNR, a higher value denotes better reconstruction quality. In this instance, the PSNR values vary from approximately 25 to 43, signifying a good reconstruction quality compared to typical values reported in existing literature for similar tasks. In the realm of image interpolation, existing literature suggests that a PSNR of 30 dB or higher is often considered good, while a PSNR of 40 dB or higher is deemed excellent [78].

These metrics indicate that the interpolation method effectively preserves structural similarity with minimal distortion compared to real intermediate slices, validating its accuracy and efficacy in this context.

Regarding the U-Net-based method, employing different loss functions during training resulted in varying performance in the reconstruction of intermediate CT slices.

- **U-Net with MSE Loss:**

This configuration yielded a very high SSIM mean value (0.994 ± 0.0064) and the highest PSNR (42.787 ± 4.52), indicating excellent reconstruction quality in terms of structural similarity and fidelity to the original image. The minimum SSIM value observed with this loss function was 0.965, and the maximum was 0.999 and, for PSNR, the minimum value was 29.645 and the maximum was 49.193. The FID score for this configuration was 6.5322. These results suggest that the MSE Loss function provides a very high-quality reconstruction with minimal distortion.

- **U-Net with SSIM Loss:**

Alternatively, using the SSIM loss function led to a slightly lower SSIM value (0.985 ± 0.0134) compared to MSE Loss and Combined Loss, but it still remained very high. The PSNR for this loss function was 39.547 ± 4.15 , which was also lower than that obtained with MSE Loss and Combined Loss. The minimum and maximum SSIM values for were 0.949 and 0.9985, respectively, while the PSNR ranged from 26.458 to 46.166. The FID score obtained was 10.828. This indicates that while SSIM Loss focuses more on specific perceptual aspects, it may not minimize global errors as effectively as MSE Loss.

- **U-Net with Combined Loss:**

Finally, with the Combined loss function, the U-Net model achieved an SSIM of 0.995 ± 0.0065 , which was comparable to that achieved with MSE Loss, and a PSNR of 42.268 ± 4.40 , slightly lower than that with MSE Loss. The minimum SSIM observed was 0.965, and the maximum was 0.999 and, for PSNR, the minimum value was 29.668 and the maximum

was 48.311. The FID score obtained was 10.166. These results indicate that the combination of losses effectively balances the maintenance of structural similarity and the reduction of global distortions.

These results highlight the effectiveness of the different loss functions in achieving high-quality reconstructions, each offering different strengths. MSE Loss provided the highest structural similarity and fidelity, as evidenced by the highest SSIM and PSNR values and the lowest FID score, indicating the best overall balance of image quality metrics. SSIM Loss, while focusing more on perceptual aspects, showed slightly lower overall metrics and a higher FID score, suggesting less effectiveness in minimizing global errors. Both aspects effectively, achieving high-quality reconstructions across all evaluated metrics, though with a higher FID score compared to MSE Loss.

Comparing the interpolation method with U-Net-based approaches reveals discernible distinctions in performance metrics. The quantitative results clearly demonstrate the superiority of U-Net over the interpolation method.

While the interpolation method demonstrates robustness with respect to SSIM and moderate PSNR, the U-Net models, regardless of the loss function used but especially when optimized with MSE loss, outperforms the interpolation in terms of both paired metrics.

This superiority underscores that the learning-based approach with U-Net effectively captures finer details and structural similarities more accurately than interpolation alone.

Furthermore, the lower FID score of U-Net models trained with MSE loss suggest higher similarity between generated and real images, highlighting the model's capacity to produce realistic reconstructions. However, it should be noted that FID scores also indicate U-Net methods (with SSIM and Combined loss functions) may slightly struggle with overall distribution fidelity compared to the interpolation method.

Next, for a qualitative analysis, two visual examples of the results from both implemented methods of Intermediate CT Reconstruction are presented: Figure 4.7 showcases the interpolation method, while Figure 4.8 illustrates the CNN-based method with U-Net trained using the MSE loss function.

These examples are juxtaposed with the provided input slices (the lower 1 and the upper 2) at the top, and the real intermediate slice (the target) is shown at the bottom left. The reconstruction output of each method is depicted in the bottom right image for comparison.

Upon analysis of the example, the interpolation method outcome appears remarkably accurate, which could be indicating the effectiveness of the employed method in this instance. The interpolated slice closely resembles the actual slice visually, maintaining consistent internal structures and contours, albeit with minor differences in finer details.

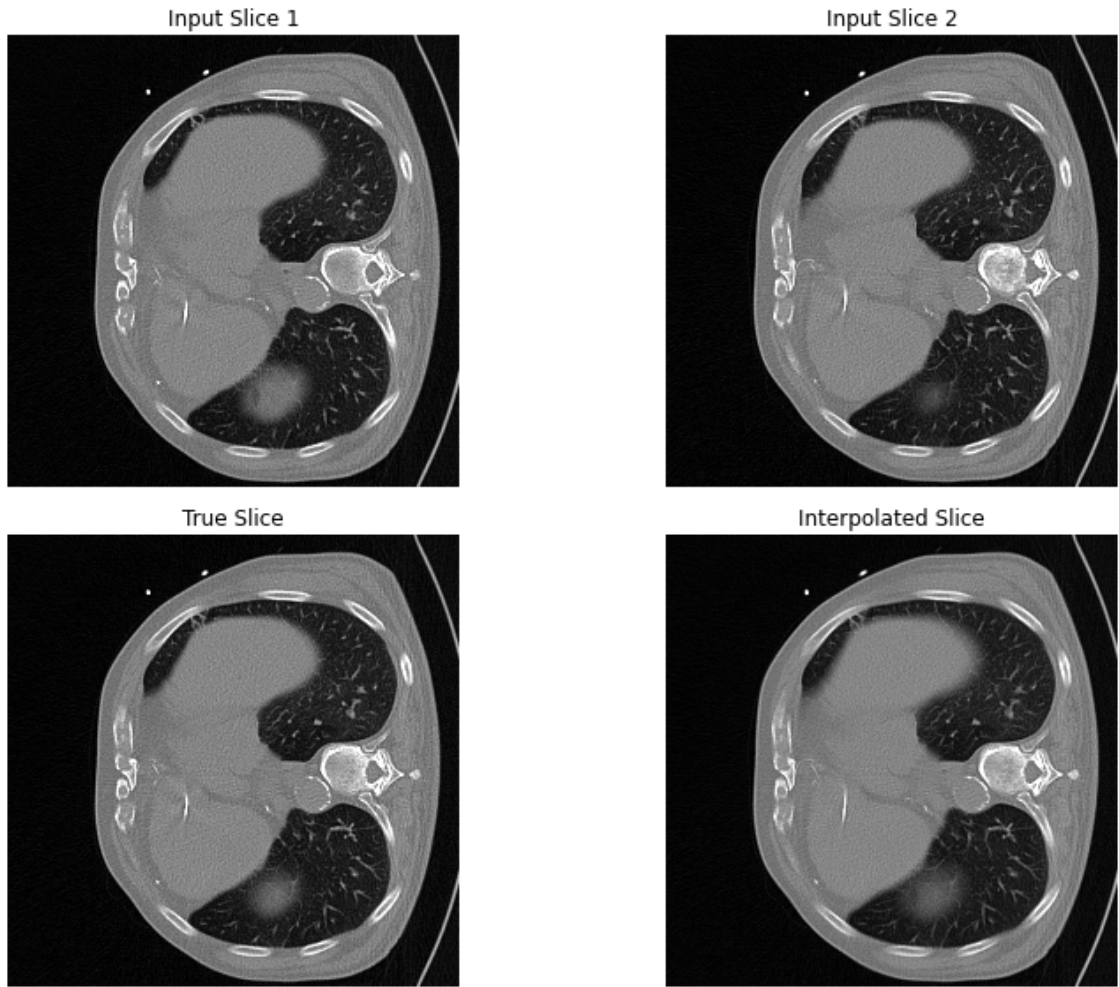


Figure 4.7: Result example of CT reconstruction using the interpolation method described. Input slices at the top, real intermediate slice at the bottom left and interpolated slice at the bottom right.

On its turn, the reconstructed image from the CNN method exhibits similarity to the target slice in terms of pulmonary structures and surrounding tissues, indicating that the U-Net model accurately captured many details present in the target slice. However, there are some subtle differences between the target slice and the reconstructed image, since the reconstructed image appears slightly less crisp than the target slice, possibly due to variations in pixel intensity or minor inaccuracies in the model's reconstruction. Overall, it seems that the U-Net model performed well in reconstructing the intermediate slice.

In Figure 4.9, side by side, the results of reconstructing the intermediate slice using both the interpolation method and the U-Net model, trained with the MSE loss function, are presented, allowing a visual comparison. This example illustrates the same slice reconstructed by both of the methods. The true slice serves as the reference image that both methods aim to reconstruct. It displays the actual internal structures and details that the reconstructed images should replicate as closely as possible.

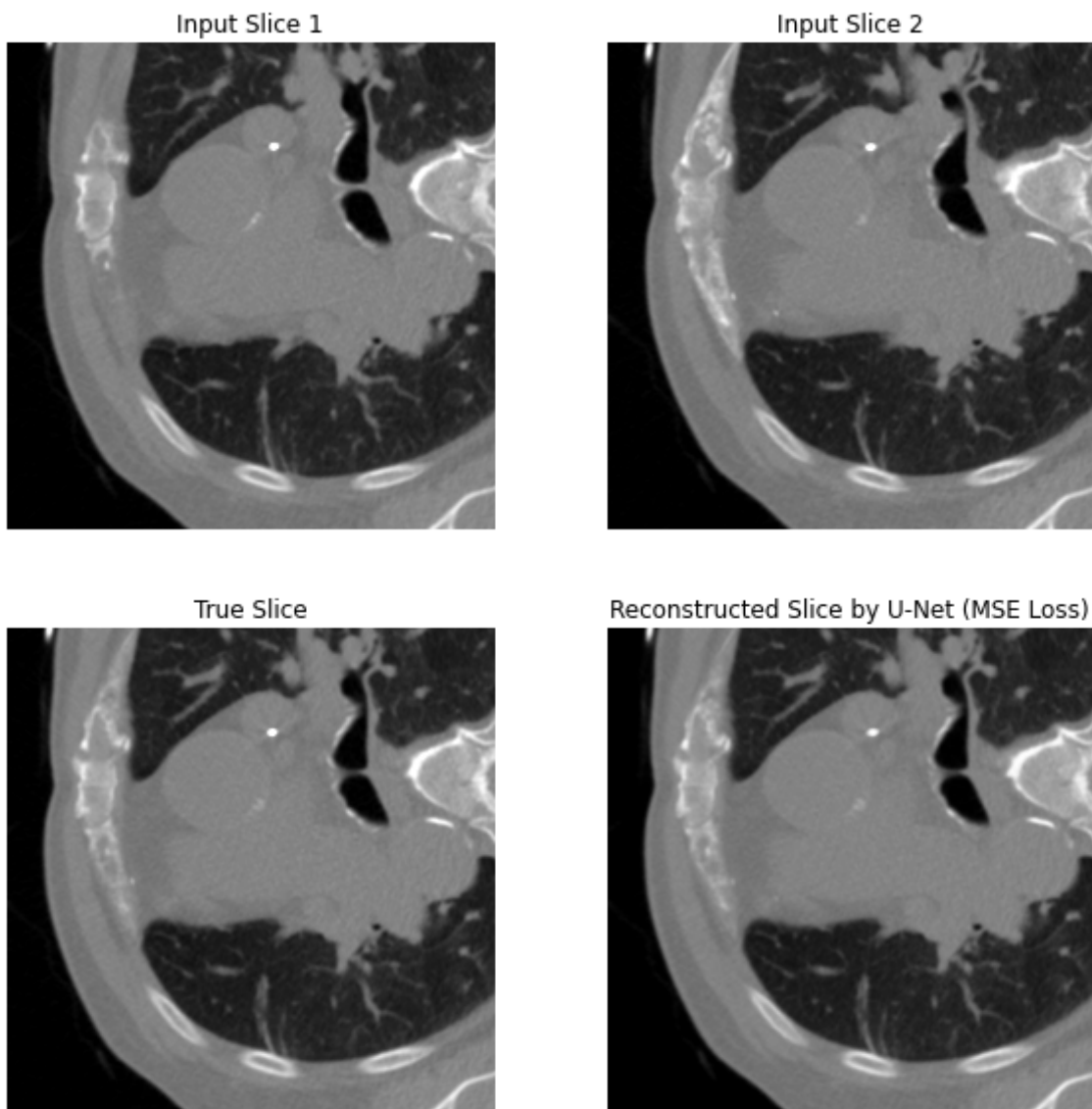


Figure 4.8: Result of model CT reconstruction using the MSE loss function: Input slices at the top, real intermediate slice at the bottom left and reconstructed slice at the bottom right.

The analysis of the figure indicates that both the interpolation method and the U-Net model succeeded in reconstructing the intermediate slice from the upper and lower slices, serving the same purpose. Nevertheless, the visual comparison between both methods' results also reveals distinct differences in the quality of reconstructed CT slices.

The interpolated slices maintain a general consistency in internal structures and contours, resembling the true slice. However, due to its straightforward averaging approach, the outcome appears smoother and can lack specific features revealing noticeable differences in finer details. This method, while effective for quick visualization, results in less sharp images and may not accurately render subtleties in tissue textures and boundaries.

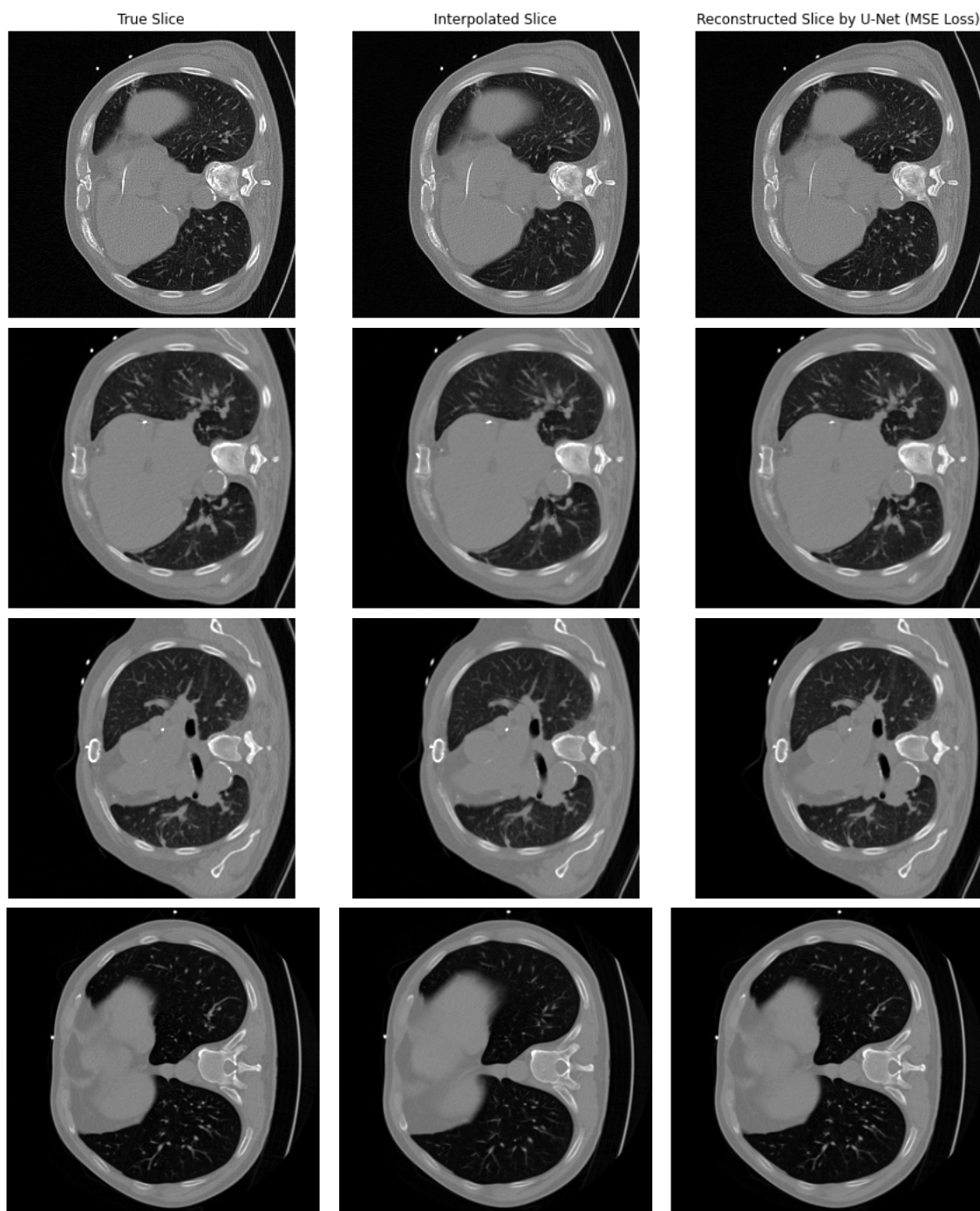


Figure 4.9: Results of reconstructing the intermediate slice using the interpolation method (middle) and the U-Net model using the MSE loss function (right), in comparison to the real target slice.

Conversely, the slice reconstructed by the U-Net model exhibits more visible details, approaching the quality of the true slice slightly more. This suggests that the U-Net model demonstrates superior performance in capturing intricate features and contextual information compared to the interpolation method.

Overall, both methods (interpolation and U-Net) performed positively in this intermediate CT Slice reconstruction task. However, while interpolation is an effective method, it seems it couldn't capture the sharpness and finer details present in the original slice as comprehensively as precise as the U-Net model. The U-Net model, on the other hand, appears to have outperformed the interpolation-based method, providing superior clarity, structural accuracy, and overall reconstruction quality, making it more suitable for medical imaging tasks, indicating that the model might be a more effective choice for CT slice reconstruction tasks.

4.5 Summary

In this chapter, two distinct methods were introduced as preliminary work towards SR in cardiac CT imaging: an interpolation-based method and a U-Net CNN-based method. Both approaches aimed to reconstruct intermediate slices from original CT scans, by utilizing the adjacent inferior and superior slices as input, allowing direct comparison with real intermediate slices to validate their effectiveness.

The interpolation-based method leveraged neighboring slices to generate intermediate slices using a averaging technique. The U-Net CNN-based method utilized a learning-based approach to achieve the same objective, with three different loss functions (MSE, SSIM, and Combined Loss) for evaluate for their impact on performance. The experiments demonstrated that both interpolation and CNN-based methods can effectively reconstruct intermediate slices in cardiac CT imaging. However, the U-Net model significantly outperformed the interpolation method, especially when using MSE Loss, which provided the highest quality reconstructions according to SSIM, PSNR, and FID metrics. The U-Net model demonstrated superior ability to preserve structural details, reduce distortions, and produce high-fidelity reconstructions, making it a more effective choice for CT slice reconstruction tasks.

In summary, this chapter marks significant achievements. The focus on reconstructing existing CT slices at intermediate positions, where ground truth is available, is imperative for affirming the accuracy and dependability of the methods employed. This groundwork before moving to the more complex task of reconstructing super-resolved CT images, which will be the focal point of the subsequent chapter, forms the foundation of our initial approach to super-resolution in cardiac CT imaging, establishing a pathway for continuous improvements. Moving forward, the next chapter will build upon these achievements by further integrating CTCA data to achieve super-resolved CT reconstructions.

Chapter 5

CT Reconstruction with CTCA Integration

This chapter builds upon the achievements from the preceding chapter by advancing the integration of CTCA data to achieve super-resolved cardiac CT reconstructions. It delves into an advanced segment of the research focused on reconstructing CT scans in previously non-existent z positions. Specifically, the aim is to reconstruct CT slices at the positions where corresponding CTCA data are available. By reconstructing CT scans at CTCA-derived positions, we seek to enhance the axial resolution of the CT data up to match the resolution level of the CTCA data, as represented in Figure 5.1. This advancement is pivotal in addressing the overarching motivation of this dissertation.

To effectively combine information from both modalities, it was necessary to address the significant disparities in imaging spatial resolutions and slice misalignments between CT and CTCA. Therefore, this chapter begins by contemplating this limitation, detailing a tailored alignment strategy designed for this purpose.

In this context, various methods were implemented, tested, and refined. The process began with the adaptation and extension of the two methodologies discussed in Chapter 4, the interpolation-based and the CNN-based techniques, to reconstruct CT scans at corresponding CTCA positions, obtaining in this way the super-resolved CT images. The outcomes of these methods were subsequently employed to test autoencoders that combine the corresponding CTCA data to enhance CT reconstruction. Finally, an adaptation of a connected autoencoders approach aimed at image separation with side information was designed and tailored to address the specific challenges of this study, transitioning to handle volumetric data.

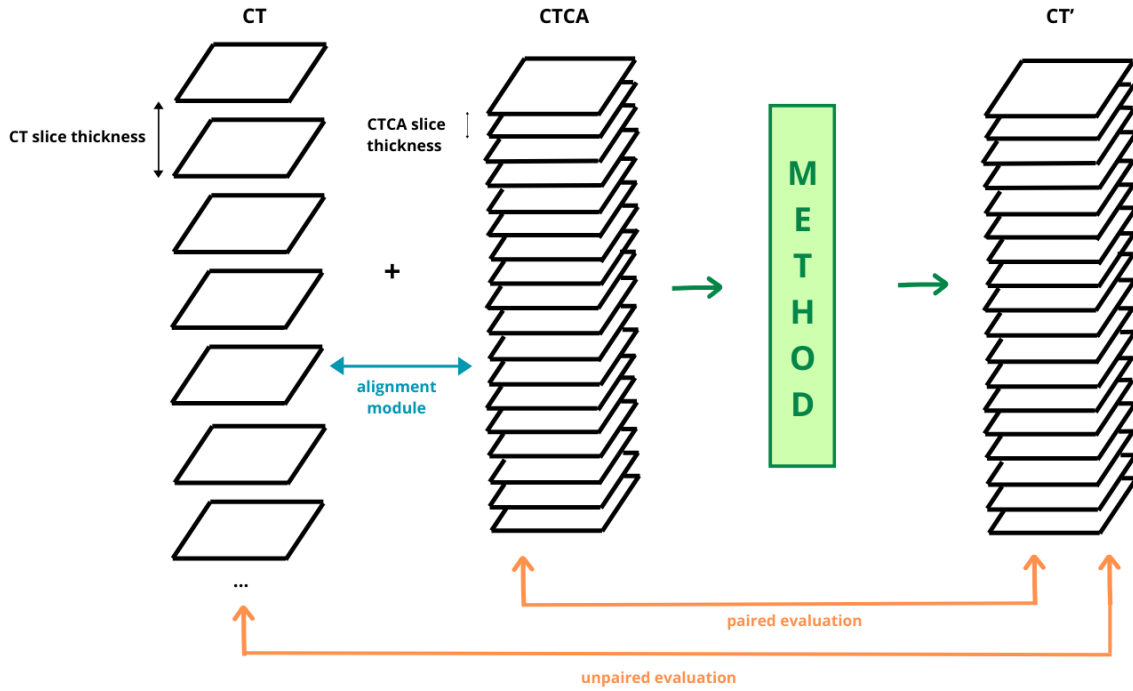


Figure 5.1: General Schematic Diagram representing the objective of the SR CT Reconstruction.

5.1 Methodology

In this section, we present all the various methods planned, constructed, and implemented within this chapter to contribute towards achieving super-resolution of CT images. The methods are described in ascending order of complexity, starting with their primary envisioned objectives. Detailed explanations of their preparation, construction, and implementation are provided.

5.1.1 Data

In this chapter, the methods implemented and tested utilized the same CHVNGE dataset as presented earlier in 4.1. The dataset was divided into 5 folds for cross-validation, each containing approximately 40 patients, as previously done. The preprocessing applied to the data also remained unchanged.

However, unlike the methods in the previous chapter that exclusively worked with CT data, this chapter incorporates CTCA data. For each patient, multiple scans of both CT and CTCA are available. The CT scans have a standard axial resolution of 1.5 mm and, in contrast, the CTCA scans vary both within the same patient and between different patients in terms of axial and spatial resolution.

Therefore, it's important to note that there are differences in spatial resolutions between CT

images and their respective CTCA images across all axis. Despite both scans possess equal dimensions of 512 by 512 pixels, a discernible mismatch was observed in the represented area, leading to non-corresponding xy positions between the two modalities, each potentially having differing fields of view (FOVs). Furthermore, due to some CTCA scans being acquired within a significantly more restricted z-axis window compared to the CT scans, a discrepancy in z positions was also evident, with the z positions of the CT not aligning with those of the CTCA.

To address this issue, the initial step in data processing involved restricting the analysis to the coinciding z-axis region between the CT and CTCA scans. Specifically, the region of interest within the CT scan was confined exclusively to the slices that corresponded to the region captured by the CTCA, i.e., the CT slices whose positions lay between the minimum and maximum z positions of the CTCA. Subsequent to this preliminary step, an algorithm was implemented to ensure precise spatial alignment between the scans from the two different modalities, as represented in Figure 5.2.

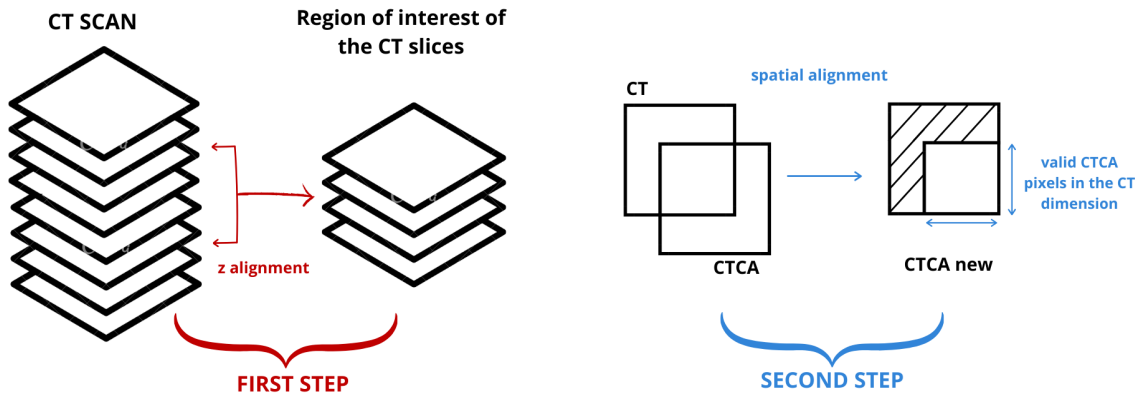


Figure 5.2: Scheme of the alignment process between CT and CTCA.

The spatial alignment algorithm obtains, for each CT slice, the nearest corresponding slice in the CTCA image. Subsequently, the spatial boundaries of the CT slice are determined to create a new space for interpolation. This space consists of 512 uniformly spaced points within these boundaries, maintaining the resolution and dimensions of the original image. Finally, a new image is interpolated within this new space, thereby obtaining a new CTCA in the same coordinates as the CT, aligning spatially with it. This process ensures spatial alignment between the CT and CTCA images for comparative analysis and allowing the CTCA incorporation into the models. The example in Figure 5.3 illustrates the outcome of aligning a CT slice with a slice from the CTCA.

5.1.2 Interpolation-Based CT Reconstruction

For the application of this interpolation method, which aims to reconstruct CT slices at previously unavailable z positions, the previously presented interpolation method 4.2.1 served as the foundation, but with a different evaluation scenario.

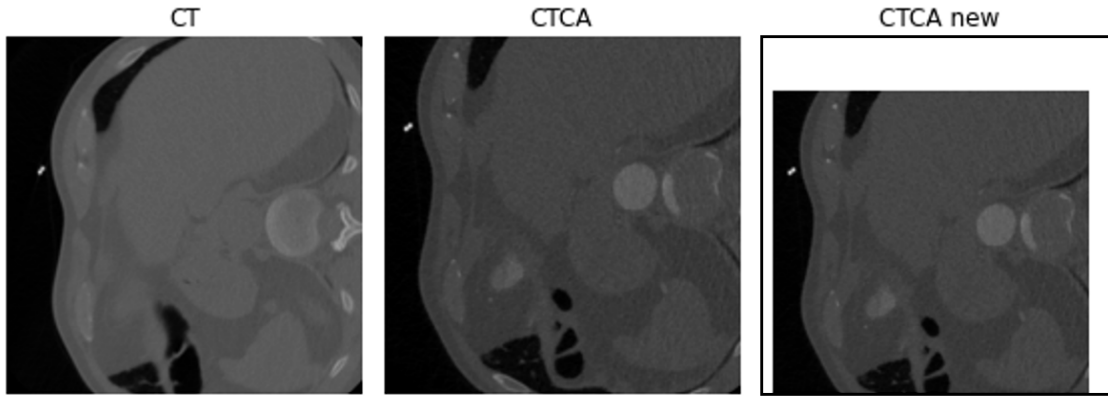


Figure 5.3: Visual result of the alignment step.

In this approach, the interpolation algorithm used weighted averaging to reconstruct a slice, i.e. in a z_3 position, positioned between two input slices, in z_1 and z_2 , representing the lower and upper bounds around the target reconstructed slice in the output. The weighted averaging interpolation technique assigns weights by taking into account the distances of z_1 and z_2 from z_3 along the z -axis. These distances play a pivotal role in determining the influence each slice holds in the output reconstruction. Notably, slices closer to the intended reconstructed slice are assigned higher weights, exerting a more substantial impact on the final reconstructed slice.

Previously, the method was evaluated by reconstructing equidistant intermediate slices, assigning equal weights to each slice during the weighted averaging interpolation technique. However, the current objective is to reconstruct the CT slice at positions where the corresponding CTCA slice is available, focusing on reconstructing slices at non-equidistant positions relative to the two input slices.

This approach enables the interpolation of slices absent in the original CT data, corresponding to positions from the higher-resolution CTCA, thereby creating super-resolved CT images.

Explicitly, for each patient, the interpolation process was performed while considering the CTCA scan acquired at a higher axial resolution. Thus, the z positions from the CT and CTCA slices were compared and, by selecting the first two slices of the CT scan, an interval delimited by their z positions was created and used to determine the number of slices from the CTCA scan that are contained in that z -axis range, retaining their corresponding z positions. In this way, it became possible to compute the distance of each of these slices from the initial CT slice position, since it corresponds to the initial slice of the CTCA.

This approach allowed the CT interpolation at z positions corresponding to those existing in the CTCA scan, so the interpolation conducted between the two input slices used weighted averages based on the distances from the input slices. Figure 5.4 illustrates a schematic representation depicting the overall behavior of this method, serving as an example for the initial iteration.

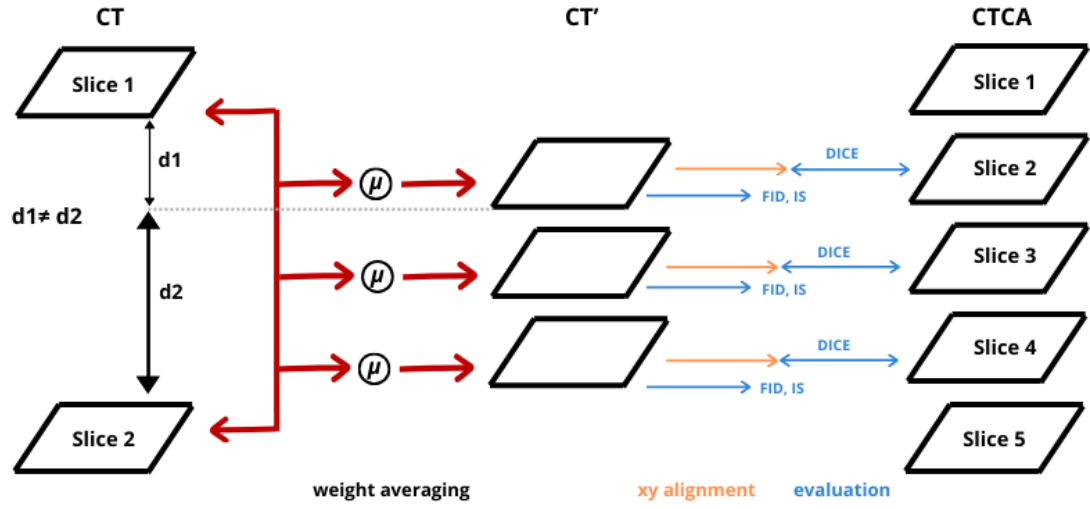


Figure 5.4: Scheme of the Interpolation-Based CT Reconstruction.

5.1.3 CNN-Based CT Reconstruction

While the previous U-Net implemented in 4.2.2 demonstrates superior performance for CT reconstruction, it has a limitation: it can only reconstruct CT slices at the intermediate position between the input slices. This restricts its ability to reconstruct CT slices at specific positions, such as those corresponding to positions from the higher-resolution CTCA, which was addressed in the previous evaluation scenario of the interpolation method.

To overcome this limitation, the U-Net represented in Figure 4.6 was adapted to accept a single CT slice as input and reconstruct that same slice as output. The architecture of the model and the training process were preserved, using MSE as the loss function.

Next, the model was tested by inputting a single CT slice into the encoder to extract its feature representation. The subsequent CT slice was then fed into the encoder to obtain its respective feature representation as well. A weighted average of these two feature representations was computed, considering the distance to the target slice position (the z position corresponding to the CTCA slices), similar to the approach previously employed in the interpolation technique, as illustrated in Figure 5.5.

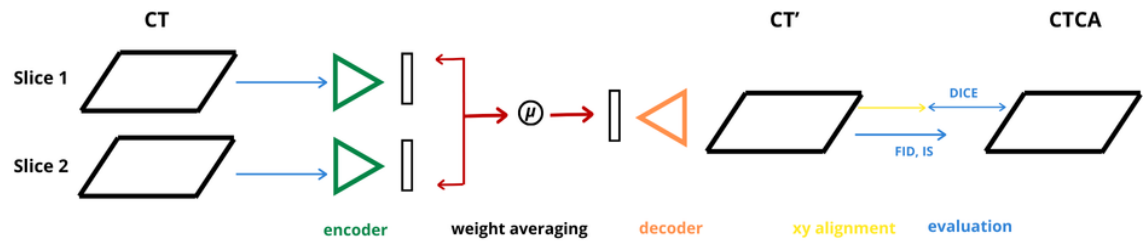


Figure 5.5: Strategy for reconstructing CT using the U-Net model.

By applying this strategy, the model should be able to generate feature representations that adapt to the specific position of the desired slice, thereby improving its capability to reconstruct

slices at various positions within the CTCA scan, much like the interpolation method used in the second evaluation scenario.

5.1.4 Autoencoder with CTCA information

To incorporate auxiliary information from the corresponding CTCA slice and improve the reconstruction of the CT slice at the consonant position, a new model was developed. This model retains the encoder and decoder architecture of the previous U-Net but removes the skip connections. These connections, typically linking encoder layers to corresponding decoder layers, were removed to allow for the independent utilization of encoders and decoders within the overall network structure, completely separating the encoder from the decoder. In this model, two encoder pathways with identical structures are instantiated, alongside a single decoder (Figure 5.6).

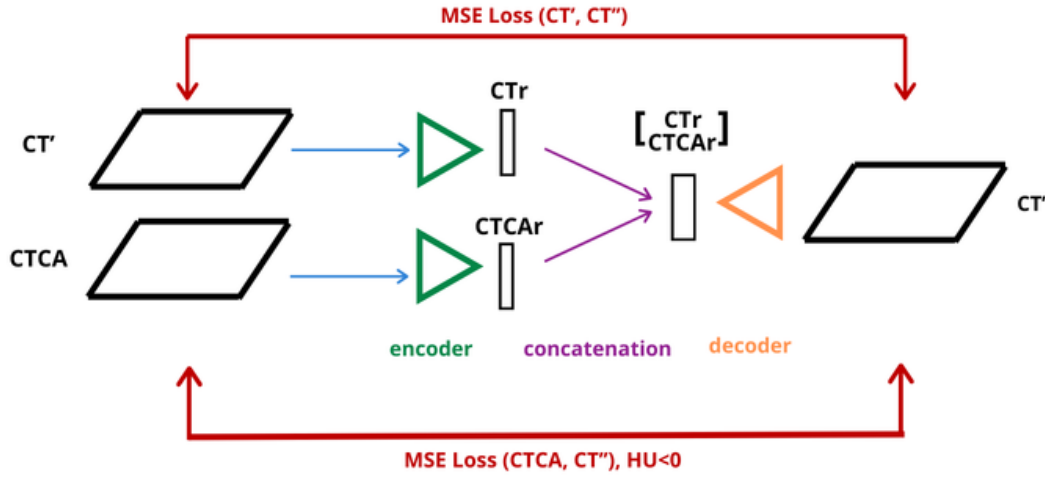


Figure 5.6: Model for enhanced CT slice reconstruction using auxiliary information from CTCA slices.

This model was trained and tested using the results of the two previous methods separately. Specifically, the input data first included the results from the Interpolation method and the corresponding CTCA slice at that position for training and testing the model. Additionally, the input data then comprised the predictions from the U-Net with weighted average in the feature representations method and the corresponding CTCA slice at that position.

The CT slice at the CTCA position is input into one encoder, while the CTCA slice is input into the other encoder, to obtain their respective feature representations. The outputs from both encoders are concatenated and fed into the decoder. The goal is to reconstruct the CT slice and assess whether incorporating information from the CTCA improves the reconstruction quality.

For data preparation, input data includes the predictions of the CT slice at the desired position and the CTCA slice from the same position.

The model was trained using the MSE loss function calculated between the model output and the input CT slice reconstructed. Additionally, the MSE loss was also calculated between the model output and the input CTCA slice. However, in this case, it was necessary to ensure that the contrast present in the CTCA image did not influence the reconstruction of the CT.

Thus, an investigation was conducted to determine the optimal threshold beyond which it would be feasible to compute the loss, aiming to extract relevant information from the CTCA image while excluding its contrast. This involved identifying pixel intensities that do not incorporate contrast.

According to the literature [79], the typical range of Hounsfield Units (HU) found in a CT scan for the human body varies from -1000 HU for air, between -100 and 100 HU for most tissues, between 100 to 500 HU for ionized contrast (depending on the administered concentration), above 1000 HU for dense bone, i.e., cortical bone, and values exceeding 3000 HU for metallic materials like implants, as shown in Figure 5.7. Furthermore, to mitigate the impact of contrast presence, different thresholds were tested on both images: below 0 HU, preserving HU values lower than 0 present in the images, and above 1000 HU, conserving values greater than 1000 HU in the images. However, given that cortical bone is not predominant in the cardiac region of the human body, applying the threshold above 1000 HU proved ineffective. Instead, the decision was made to apply the threshold below 0 HU, allowing predominantly fat tissues and air to remain visible.

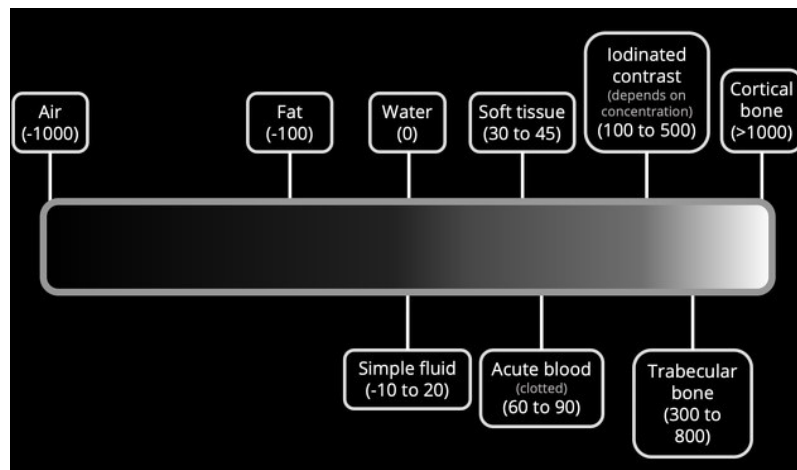


Figure 5.7: Hounsfield scale with simplified reference values for different structures [79].

Therefore, the MSE loss between the model output and the input CTCA slice was calculated exclusively for pixels with $HU < 0$. All other training parameters and strategies remained the same, except for the batch size, which was reduced to 4.

By implementing this approach, the model aims to determine if integrating information from the CTCA can enhance the reconstruction accuracy of the CT slice.

5.1.5 2D Connected Autoencoders

In a subsequent phase, the integration of information from CTCA in a more complete way was considered. The study titled "Image Separation With Side Information: A Connected Auto-Encoders Based Approach" [80] inspired a new strategy to address the problem's motivation.

The paper proposes a novel neural network model for the effective separation of mixed X-ray images, specifically in double-sided paintings. X-radiography of double-sided paintings results in a mixed X-ray image, and this paper addresses the problem of separating this mixed image. Thus, the model proposed by Wei Pu et al. employs connected autoencoders (CAEs) to separate the mixed X-ray image into two simulated X-ray images, each corresponding to one side of the painting. The architecture of the AEs is grounded in iterative shrinkage convolutional learning algorithms (CLISTA). In this model, the encoders extract sparse codes from the visible (RGB) images of both the front and back of the paintings, as well as from the mixed X-ray image. The decoders, in turn, are responsible for reproducing the original RGB images and the mixed X-ray image.

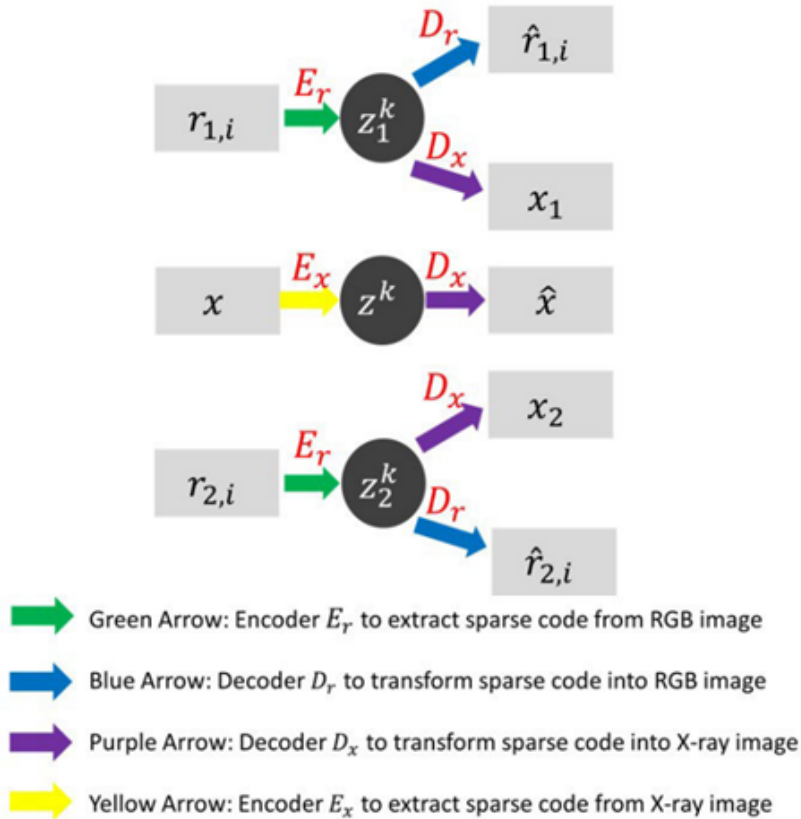


Figure 5.8: General structure of the CAEs model from [80].

Figure 5.8 illustrates the general structure of the CAEs model used for separating mixed X-ray

images of double-sided paintings. In this model, x represents the mixed X-ray image patch, while x_1 and x_2 represent the individual X-ray image patches. The notations $r_{1,i}$ and $r_{2,i}$ for $i = 1, 2, 3$ denote the red, green, and blue channel patches of the RGB images from the front and rear sides of the painting, respectively.

The encoded sparse representations are denoted by z_1 and z_2 , which correspond to the image patches on the front and rear sides of the painting. The reconstructed images are denoted by $\hat{x}$, $\hat{r}_{1,i}$, and $\hat{r}_{2,i}$, which represent the mixed X-ray image and the regenerated RGB images of the front and rear paintings for the i -th channel, respectively.

During training, the encoding and decoding networks are trained in a fully self-supervised manner using a joint loss function that integrates the various constituent networks. The loss functions employed in the study are designed to ensure the effective separation of the mixed X-ray images. These include:

1. **Reconstruction Loss:** This loss ensures that the connected AE maps the patches of the RGB images and the mixed X-ray image back to themselves, promoting the fidelity of the reconstruction of the original images.
2. **Sparsity Loss:** This loss encourages the sparsity of the underlying representations, incentivizing the use of only a reduced number of non-zero elements in the data representation.
3. **Constraint Loss:** This loss ensures that the mixture of the sparse feature maps adheres to the physical constraint that the mixed X-ray image must be the combination of the two separated images, respecting the Beer-Lambert law for X-ray attenuation. This loss is specific to x-ray images.

The method has been tested on images of the wing panels of the Altarpiece of Ghent, demonstrating that the proposed approach outperforms other X-ray image separation methods for art investigation applications. The results indicate a significant improvement compared to state-of-the-art image separation methods, particularly in terms of preserving informative details about the surface and subsurface characteristics of the paint layers on a specific side of the painting.

Similar to our scenario, we are presented with two representations of the same entity, where one of these representations is a combination or mixture of the two. In the article's context, this mixture comprises X-ray images taken from each side of the painting, along with an X-ray image of their overlay. In our case, we deal with a CT and CTCA scans, where the CTCA can be seen as a combination of the CT scan and a virtual contrast information scan. This comparison highlights the parallel between the challenges faced in both situations, providing a clearer understanding of the problem context.

Therefore, drawing inspiration from the CAEs approach outlined in the article, the model aimeded to implement comprised two encoders (E1 and E2). Encoder E1, which would receive

CT data and extract their feature representation, and encoder E2, which would perform the same process with CTCA data. Complementing these encoders, two decoders (D1 and D2) were envisioned: decoder D1 would receive the concatenated outputs of both encoders and produce two outputs (the CT reconstruction and the contrast reconstruction). On the other hand, D2 would take the output of encoder E2, i.e., the feature representations of CTCA data, and generate the CTCA reconstruction as its output. The schematic diagram illustrating the envisioned architecture for the model is depicted in the Figure 5.9.

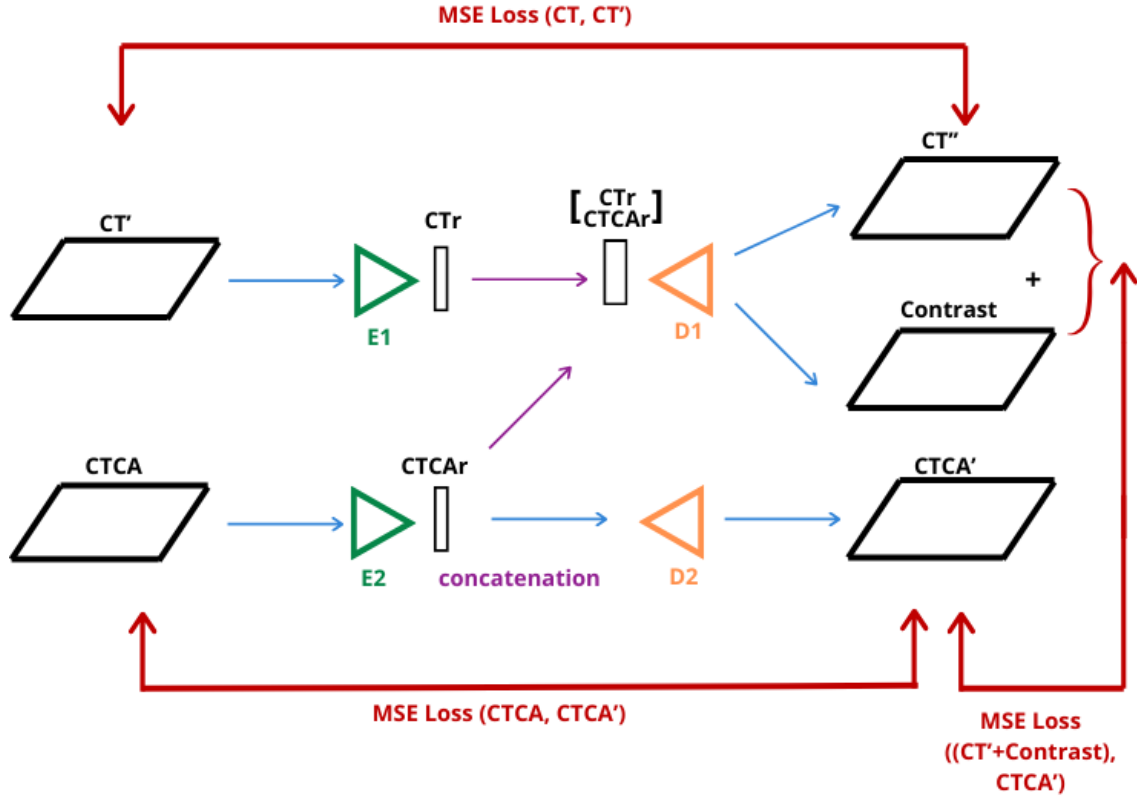


Figure 5.9: General architectural framework of the 2D connected AEs.

The model was implemented in a 2D scenario, receiving as input the output from a previous step—specifically, a CT slice reconstructed at a CTCA position from the U-Net with weighted average in the feature representations (CNN-Based CT Reconstruction), along with its corresponding CTCA. The CT slice is fed into encoder E1, while the CTCA slice is fed into encoder E2, both working to encode the images into feature representations. The combined output from both encoders is given to decoder D1 to incorporate information from the CTCA, as this decoder aims to reconstruct the input CT slice and a slice representing the contrast present in the CTCA. Simultaneously, the output from encoder E2 is provided as input to decoder D2, which reconstructs the input CTCA slice.

The architecture of this model adheres to the structure of the previous model, maintaining the encoder structures while modifying the decoder to provide two outputs (reconstruction of the

CT slice and the contrast). Additionally, a new simple decoder was instantiated to receive the feature representations of the CTCA slice and reconstruct it at its output. Similar to the earlier presented method, the skip connections to fully segregate all the encoders from the decoders were also eliminated.

During training, the encoders and decoders are trained using the MSE loss to connect the various constituent networks, drawing inspiration from the training approach in the referenced article. In each iteration, three different losses were calculated, as represented in the figure:

1. The loss between the reconstructed CT slice output from Decoder D1 and the original input CT slice.
2. The loss between the output slice from Decoder D2, which reconstructs the CTCA, and the original input CTCA slice.
3. The loss between the sum of the output slices from Decoder D1 (CT reconstruction + contrast reconstruction) and the output slice from Decoder D2.

The training details are as follows: a learning rate of $1e-4$, a batch size of 4, a maximum of 300 epochs, with early stopping applied if there was no improvement after 20 epochs.

This model serves as an advanced method for integrating CTCA slice information into the reconstruction of CT slices at corresponding CTCA positions, thereby implementing principles of image separation within the realm of medical imaging. This strategy represents an intermediate step towards the application of this methodology to volumetric (3D) data.

5.1.6 3D Connected Autoencoders

Previously, all methods were treated in a 2D context, focusing exclusively on the reconstruction of individual CT slices corresponding to specific CTCA positions. This approach, while effective, is inherently limited in its scope. The transition to a 3D framework could be beneficial to enhance the accuracy and utility of the reconstructions by considering the volumetric data as a whole, rather than isolated slices. Furthermore, by extending the current 2D approach to 3D, we aim to leverage the full potential of CAEs in medical imaging, ensuring more robust reconstructions.

The model was thus implemented in a 3D scenario (Figure 5.10), increasing its complexity and requiring additional considerations for data preparation, architecture, and training. Data setup was adjusted to handle volumetric data, and the encoders and decoders were adapted for 3D data. These adjustments were essential for maintaining model performance and accuracy in reconstructing and interpreting 3D CT and CTCA slices.

Data Setup

During each iteration of the model, the input data now comprises two consecutive CT slices and a variable number (N) of CTCA slices, where N represents all the CTCA slices located between

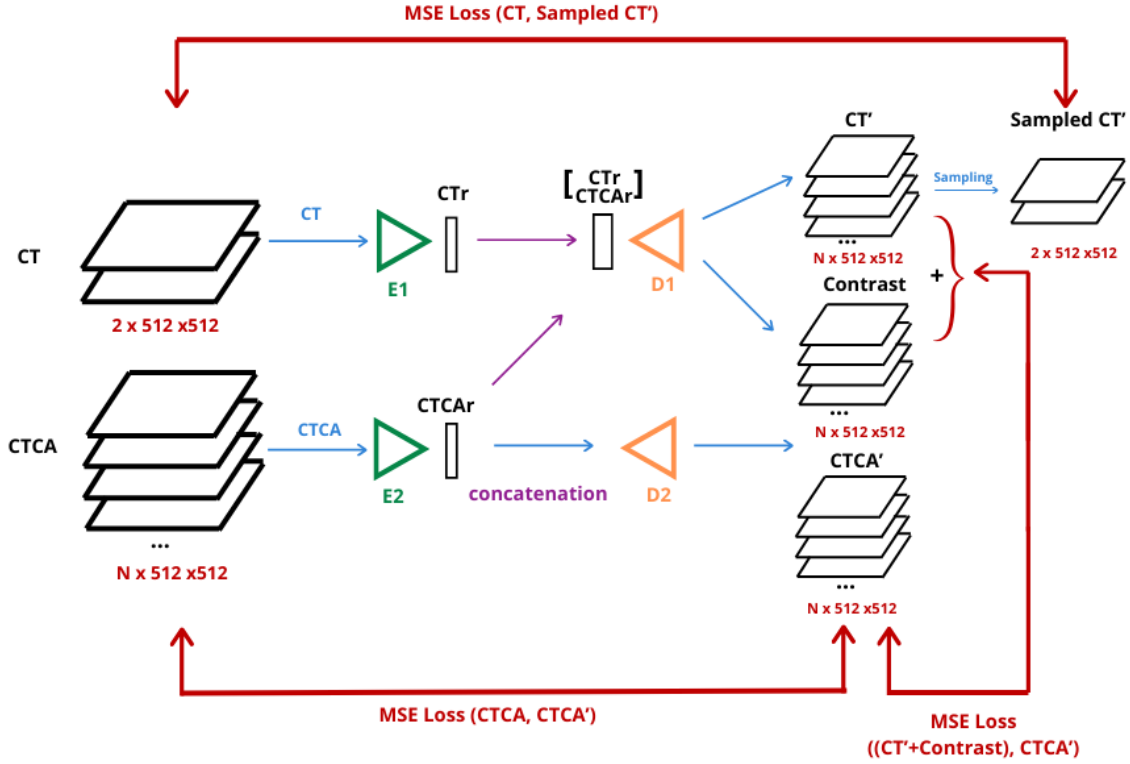


Figure 5.10: General architectural framework of the 3D connected AEs.

the z positions of the input CT slices, and also including the CTCA slices present within half a slice thickness distance from both the lower CT slice and the upper CT slice, to provide additional contextual information. The number of CTCA slices varies depending on the specific axial resolution of the CTCA scan. To enrich the dataset, data preparation involved creating combinations of all CT and CTCA scans for each patient. Alignment between the two modalities was managed according to the previously outlined procedure.

To use a batch size larger than one, it is also necessary to ensure uniformity in the number of CTCA slices across all samples within a batch. To meet this requirement, a custom batch sampler was designed and integrated into the PyTorch Dataloader framework, aimed at creating data batches wherein every constituent sample in a batch has the same number of CTCA slices, thus ensuring homogeneity within each batch (Figure 5.11).

Before implementing the batch sampler, an analysis of the number of CTCA slices across all examples was conducted to facilitate the categorization of the data into distinct index lists based on slice count. Subsequently, during the sampler iteration process, the first sample was randomly selected, and the corresponding CTCA slice count was verified. Then, until completing the batch, the remaining samples were directly selected from the designated index list with the same slice count, ensuring uniformity in CTCA slice count across all samples within the batch. This step was crucial as it ensured the consistency of batch composition with regard to CTCA slice count, allowing the use of a batch sizes exceeding one, significantly improving training speed and efficiency.

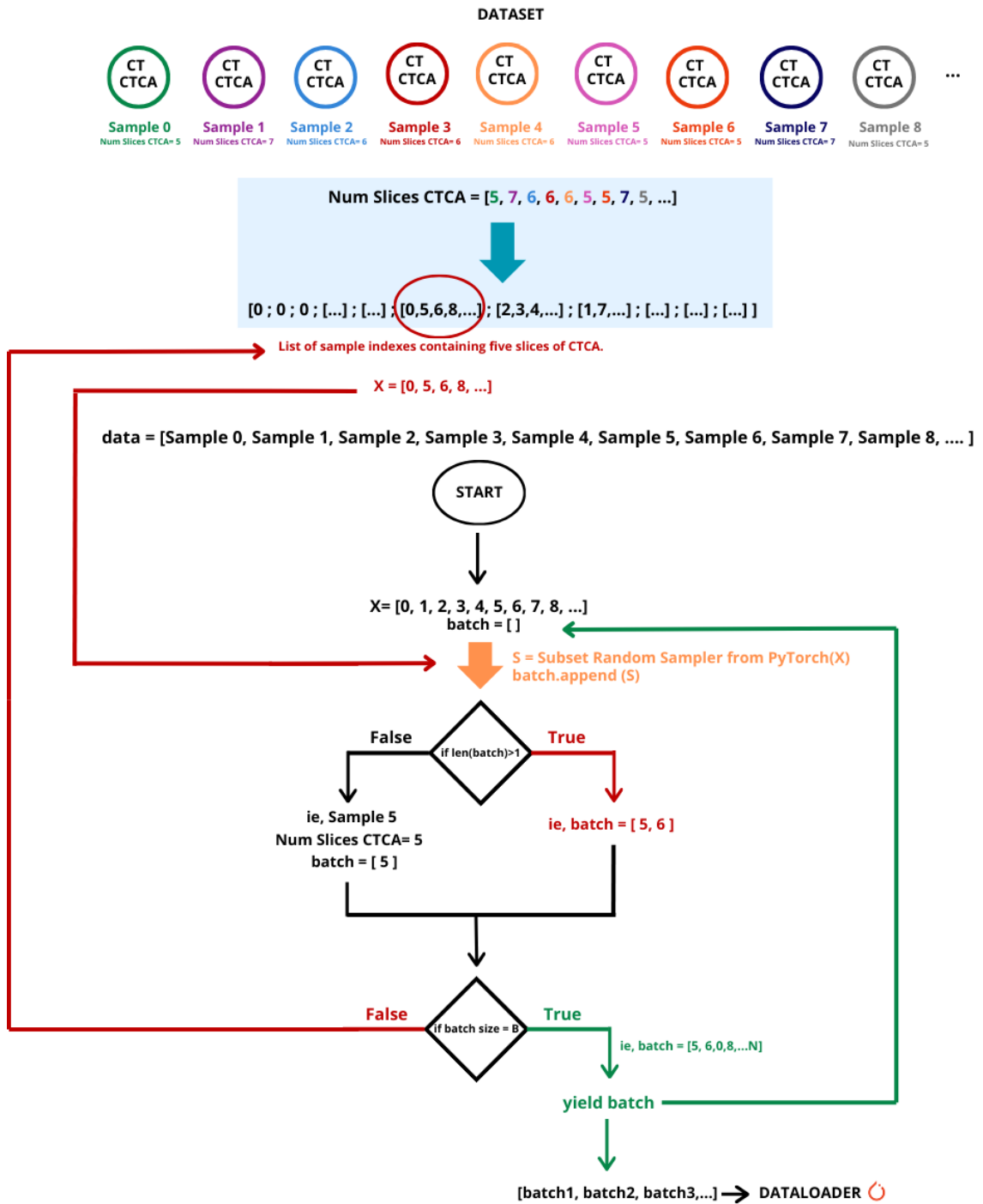


Figure 5.11: Illustrative example of how the implemented batch sampler works.

Model Architecture

Maintaining the same base architecture as the 2D scenario, the model comprises two encoders and two decoders. In this case, Encoder E1 is responsible for processing the input file containing two

consecutive CT slices, while Encoder E2 handles the N input CTCA slices. The outputs of both encoders are concatenated and fed into Decoder D1, enabling the utilization of information from both modalities. Decoder D1 aims to reconstruct the CT with the same number of slices as CTCA, as well as reconstruct the contrast present in CTCA. Decoder D2, on the other hand, only receives the output of Encoder E2 and functions to reconstruct the input CTCA.

As in previous subsections, a U-Net architecture without skip-connections was used in this approach. Then, to transition the network to a 3D scenario, $3*3*1$ convolutions followed by $1*1*3$ convolutions were employed instead of standard 3D convolutions [81]. The $3*3*1$ convolutions are applied to each slice of the 3D input volume, allowing the network to capture two-dimensional spatial patterns in each layer while maintaining the volume's dimensionality along the Z-axis. Subsequently, $1*1*3$ convolutions are employed to integrate information across different layers of the resulting volume from the previous step. This approach aims to provide 3D information in a computationally efficient way, since by using $3*3*1$ convolutions, the number of parameters required is reduced compared to standard 3D convolutions as they operate on each slice individually.

The adopted approach followed these steps:

1. Training a simple AE to receive and reconstruct 2 CT slices (Figure 5.12a).
2. Training another AE to handle and reconstruct N CTCA slices (Figure 5.12b).
3. Combining the components of these AEs from steps 1 and 2 to create the comprehensive 3D AE system, leveraging these training steps.

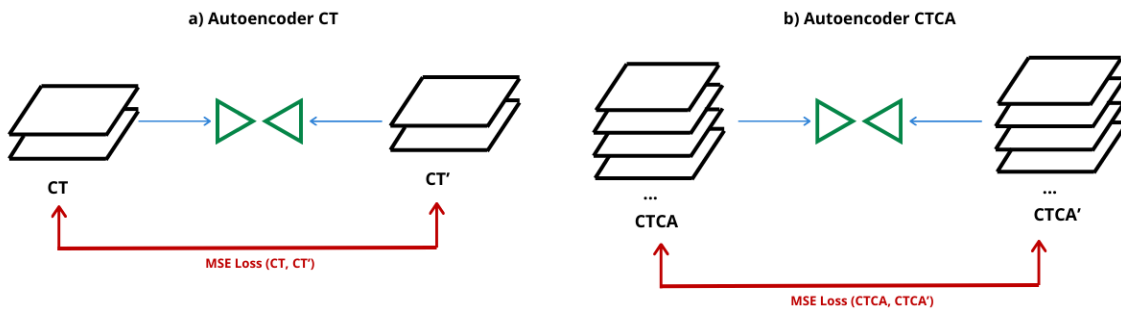


Figure 5.12: General architectural framework of: a) CT autoencoder and b) CTCA autoencoder.

The first two steps served as pretraining, with their training being useful for subsequently leveraging the weights from these CT and CTCA autoencoders to initialize the final network. This approach is particularly beneficial since 3D networks are significantly more complex and difficult to train.

In the development of the CTCA AE, additional considerations were made regarding the dimensions of the features, aiming to ensure that the network architecture could accommodate the variability in the number of slices in the input images.

Given that different CTCA datasets have varying numbers of slices, it was imperative to ensure that the neural network capable of processing CTCA images with a variable number of slices while still generating a consistent set of features in the bottleneck layer in terms of shape. This necessitates that the network operates in an agnostic or flexible manner with respect to this input dimension. The input data format is $[B, C, N, H, W]$, where B represents the batch size, C denotes the number of channels, N indicates the number of slices, and H and W represent the spatial dimensions of the image (Height and Width). Regardless of the number of slices in the input image, the network should transform the images in a way that, upon reaching the end of the encoder, the N dimension is reduced to 1, while the spatial dimensions (H and W) are halved, and the number of channels (C) is doubled along the downsampling path. If the N dimension reaches 1 before the end of the downsampling blocks, it remains constant at 1, with no further downsampling occurring in that dimension, while the reduction of H and W and the increase of C continue. This consistency at the bottleneck enables the network to uniformly process volumes of different sizes.

In the decoding phase, the process is reversed with the objective of restoring the original dimensions of the image by reducing C , increasing H and W , and increasing N until the original number of slices is reached. This ensures that the number of slices (N) in the output matches that of the input. When the dimension N reaches the original number of slices, the upsampling in this dimension ceases. This approach guarantees that the output of the decoder maintains the same dimensions as the original input, thereby preserving the consistency and accuracy of the image reconstruction.

These additional considerations adopted for the CTCA AE are essential to ensure the flexibility and robustness of the neural network when dealing with medical images with varying numbers of slices. These adjustments enable the network to produce a uniform set of features at the bottleneck and correctly restore the original image dimensions, regardless of variability in the input dimensions.

Training Details

Both CT and CTCA AEs employed the same training strategy as used previously, with a learning rate of $1e-4$. The batch size was set to 6 for the CT AE and 4 for the CTCA AE. The maximum number of epochs was 300, and training would stop if there was no improvement after 20 epochs. For the CT AE, the input data consisted of pairs of consecutive CT slices, while the CTCA AE's input data consisted of N CTCA slices. Both AEs were trained using the MSE loss between the input and output.

As mentioned earlier, the final implementation of the 3D AE utilized pre-trained CT and CTCA AEs for weight and bias initialization. Specifically, the encoder from the CT AE initialized encoder $E1$, the encoder from the CTCA AE initialized encoder $E2$, and the decoder from the CTCA AE initialized decoder $D2$. Meanwhile, decoder $D1$ was designed to integrate the concatenated outputs of both encoders to reconstruct CT images and the contrast present in CTCA.

The weights and biases of this decoder were initialized using the Xavier Uniform initialization technique [75].

Maintaining the same training logic as the 2D scenario, in each iteration, three different losses were calculated, as illustrated in Figure 5.10, albeit in a 3D scenario.

1. **Loss between the output of decoder D2 and the original CTCA input:** This loss assesses the difference between the reconstruction generated by decoder D2 and the original CTCA image.
2. **Loss between the sum of the outputs of decoder D1 and the output of decoder D2:** This metric compares the sum of the reconstructions of the CT and the contrast produced by decoder D1 with the output of decoder D2.

Both of these losses were calculated only in the region where pixels are present in both images, considering that the CTCA was previously aligned.

3. **Loss between the result of a sampling operation on the CT reconstruction output of decoder D1 and the original CT input**

To calculate this last loss, it was necessary to apply a sampling operation to the CT reconstruction output, as the output of the decoder had the same number of slices as the CTCA, making a direct comparison with the original CT, which had only 2 slices, impossible. The sampling operation aimed to adjust the output of decoder D1 to obtain a CT reconstruction with 2 slices, thus allowing the calculation of the MSE loss between this result and the original CT input.

The applied sampling operation employed weighted averaging based on the z-axis distances between the positions of the CT reconstruction slices, i.e., the CTCA slices, and the positions of the CT slices.

To elaborate, the positions of the CT and CTCA slices were loaded from text files and then, for each slice in the original CT, the absolute difference between its position and the positions of the CTCA slices was calculated. Based on these differences, weights inversely proportional to the distances were computed and normalized to sum up to 1. These weights were then applied to the CT reconstruction slices, resulting in weighted slices based on their relative proximity to the original CT slice. For each original slice of the CT, the corresponding slice in the reconstruction is obtained by the weighted sum of the CTCA slices. The final result is a reconstruction of the CT with only 2 slices, where each slice is a weighted combination of the original CTCA slices, based on their relative positions along the z-axis. This approach ensures that the final reconstruction is compatible with the dimension of the original CT for the calculation of MSE loss. The Figure 5.13 illustrates the operation of the sampling process.

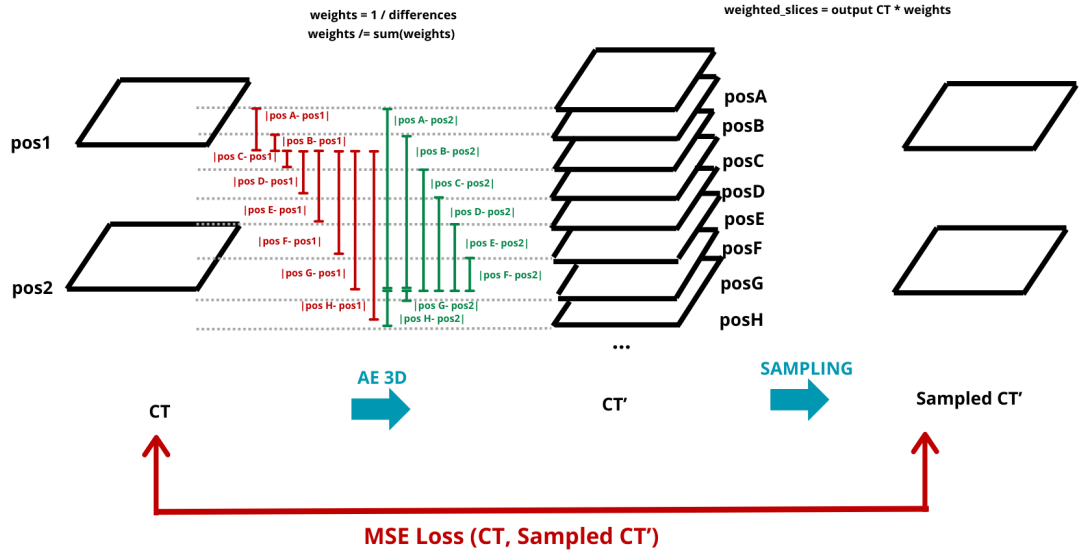


Figure 5.13: Diagram depicting the sampling process operation.

Finally, in computing all the losses, an additional term related to sparsity loss was introduced to promote the sparsity of the representations learned by the model. This was implemented with an L1 penalty applied to the intermediate activations of the model, namely the outputs of the encoders. The L1 Loss function calculates the sum of the absolute magnitudes of the activations, promoting sparsity by encouraging many values to be zero or close to zero, thus utilizing the fewest features possible.

This technique aids in improving the model's generalization, reducing the risk of overfitting by preventing the model from overly relying on specific characteristics of the training data [82]. Additionally, it can lead to greater computational efficiency, as operations can be executed more quickly and with less memory consumption.

Thus, the sparse loss equation combines the MSE loss with an L1 penalty on the intermediate activations, expressed as follows:

$$L = L_{MSE}(CTCA, CTCA') + L_{MSE}((CT' + \text{Contrast}), CTCA') + L_{MSE}(CT, \text{Sampled}CT') + \alpha(\|ct_{enc}\|_1 + \|ctca_{enc}\|_1) \quad (5.1)$$

Where L_{MSE} are the MSE losses, $\|ct_{enc}\|_1$ and $\|ctca_{enc}\|_1$ are the L1 norms of the intermediate activations of encoder E1 and E2, respectively, and α is a hyperparameter regularization factor that controls the importance of sparsity penalty relative to MSE loss. Here, the α value used was 0.001.

It is worth noting that training this model was considerably more computationally intensive compared to the previous models. Thus, this deep learning model was trained using PyTorch and 2 NVIDIA TITAN RTX GPUs. The training configuration included a batch size of 4, a learning rate of $1e-4$, and a maximum of 100 epochs, with early stopping implemented if the three losses did not improve after 20 consecutive epochs.

5.2 Evaluation

Given that, for the results in this chapter, the generated CT slices lack a direct reference in the original CT data (i.e., there is no ground truth CT available for paired evaluation), unlike the previous chapter, alternative evaluation approaches were required to assess the accuracy and reliability of the generated slices. This involved assessing the realism of the obtained results using metrics independent of the ground truth, unpaired metrics, as well as conducting an anatomical comparison between the generated CT slice and its corresponding CTCA slice.

The selected metrics assess statistical similarity, image quality, pixel-level difference, and spatial anatomy, enabling a comprehensive evaluation of the methods' ability to generate high-quality and anatomically accurate images despite the absence of ground truth.

Realism:

When assessing images without an available ground truth, unpaired metrics are commonly used aiming to evaluate their realism, specifically the FID score.

In this chapter, mirroring the evaluation of the Intermediate CT Reconstruction methods, the FID score was also employed as an evaluation metric for the tested methods. Additionally, the adjusted FID value (FID*) was calculated to refine the evaluation since it specifically considers the region of the CT slice that corresponds to the CTCA, rather than the entire CT slice. This approach ensures that the comparison focuses on the relevant anatomical structures where CT and CTCA overlap. By calculating FID* in this manner, the assessment becomes more targeted and reflective of how accurately the generated images capture the features of interest in the integrated CTCA data region.

Anatomical Comparison:

In this evaluation scenario, since the CT reconstruction was performed at the z positions where there is correspondence with slices of the CTCA image, it would be pertinent to compare both images to assess their similarity.

However, a direct anatomical comparison between the generated image and the corresponding CTCA slice is not feasible due to the presence of contrast agent in the CTCA image. To make this comparison valid, the previously mentioned threshold ($HU < 0$) was applied to both images in order to ensure that the contrast present in the CTCA image did not influence this comparison, since its application has the same purpose as previously but for a different task.

By applying the threshold, the calculation of the DSC was performed directly between the binary masks of each image, created based on this threshold. Here, pixel values in the CT and CTCA images equal or lower than this threshold were defined as True, while the rest were marked as False. Figure 5.14 illustrate an example demonstrating the implementation of thresholding on

CT and CTCA slices, where the top images shows the lower original CT slice and the corresponding CTCA slice, respectively, before thresholding and the bottom images presents the binary mask of the CT and CTCA slices after thresholding.

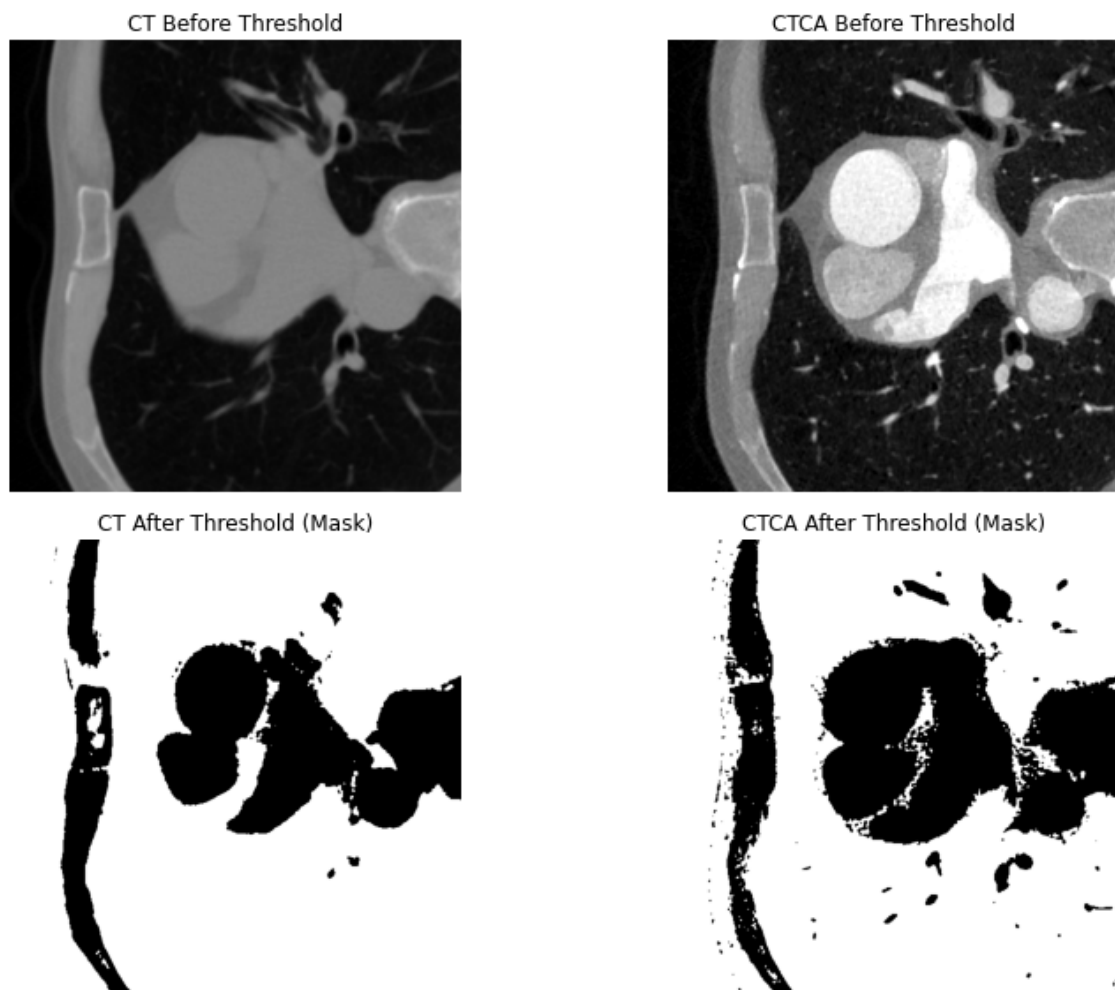


Figure 5.14: Result example of the application of the threshold, for computing the DSC.

The DSC is a statistical measure commonly utilized to assess the similarity or overlap between two sets of data. It is widely used in comparing concordance or overlap between two regions, such as two binary masks representing different anatomical structures or areas of interest in medical images. DICE coefficient ranges from 0 to 1, where a value of 0 indicates no overlap between the sets while 1 signifies complete agreement between them. Thus, a DSC closer to 1 indicates higher similarity between the compared sets, denoting greater concordance between the evaluated areas or regions.

5.3 Results and Discussion

This section presents both quantitative metrics and qualitative evaluations through visual examples for the implementation of the aforementioned methods. It is divided into three subsections, each focusing on distinct implementation objectives. Firstly, in 5.3.1, the results of the Interpolation-based and CNN-based methods are presented. These methods aim to reconstruct the CT at CTCA-derived positions, leveraging distances from the desired slice position to input slice positions. Next, in 5.3.2, results from two AEs implemented with integrated CTCA information are presented. These AEs are designed to enhance CT reconstruction using the outputs from the previous subsection. Lastly, in 5.3.3, the results from the two CAEs approaches are presented with a common goal of integrating information from CTCA and principles of image separation for CT reconstruction, both in 2D and 3D contexts. Each subsection analyzes and discusses the obtained results, including comparisons with previous methods.

5.3.1 CT Reconstruction

The initial methodologies implemented in this chapter concentrated on reconstructing CT images at positions with corresponding CTCA data. These methods involved providing the input CT data and extracting the z-axis positional information from the CTCA data to perform the reconstruction at those specified positions. Table 5.1 presents the quantitative results obtained for the metrics calculated for both methods.

Table 5.1: Results for CT Reconstruction methods (Interpolation VS CNN-Based).

Method	DSC	FID	FID*
1. Interpolation- Based	0.709 ± 0.147	7.6222	7.6222
2. CNN- Based	0.719 ± 0.121	6.8898	6.8898

The results obtained for both methods were quite similar, however, they indicate that the CNN-based method outperforms the interpolation-based method across all evaluated metrics.

Specifically, the CNN-based approach achieved a slightly higher DSC (0.719 ± 0.121) compared to the interpolation-based method (0.709 ± 0.147), indicating a marginally better overlap between the reconstructed CT and the corresponding CTCA masks. The reduced variance in the CNN-based method's DSC also suggests more consistent performance across different images. This improvement suggests that the CNN-based model is more adept at capturing and reconstructing anatomical structures. However, given the closely aligned values, it can be concluded that both methods achieved a reasonable level of accuracy and structural similarity between the reconstructed CT images and the CTCA masks, suggesting that both approaches are effective in reconstructing anatomically accurate slices.

The results from the unpaired metrics calculation are crucial in evaluating the performance of CT reconstruction at CTCA positions, as we are generating CT slices that do not exist in the original data, thus lacking a real CT reference. Therefore, these metrics provide insights into the quality and realism of the reconstructed CT slices from various perspectives.

The FID values further reinforce the CNN-based method's superiority, with a lower FID of 6.8898 compared to 7.6222 for the interpolation-based method. The lower FID indicates that the distributions of the generated images are closer in distribution to the real images, signifying higher quality and more realistic reconstructions. This suggests that the CNN-based method is better at capturing the underlying data distribution, resulting in more realistic reconstructions. Additionally, both methods yielded the same results in FID* values since the CTCA region doesn't significantly impact the difference since there are no variations within or outside the CTCA regions.

Comparing the FID values obtained from the Interpolation-Based and CNN-Based methods in this chapter with those from the previous Chapter 5.3, which served as the foundation for advancing these methods from reconstructing intermediate slices present in the original CT to reconstructing slices at corresponding CTCA positions that do not exist in the original data, we find that there were no significant differences in the values. They remained quite similar and within the same range of values, which indicates that these methods maintain their performance even for reconstructions of slices that are not originally present.

Usually, in medical contexts, a lower FID score, generally signify a closer match between the distributions of real and generated images, indicating better performance [83]. Compared to values commonly observed in the literature, the obtained metrics appear to be within the expected range [83, 84, 85].

The overall superior performance of the CNN-based method, evidenced by higher DSC scores and lower FID scores, underscores the advantages of employing deep learning techniques for CT image reconstruction, similar to the findings from the previous chapter. The U-Net model's capacity to learn intricate features and accurately represent spatial structures contributes significantly to its superior performance compared to the interpolation-based method. While simpler, the interpolation method falls short in capturing detailed anatomical nuances, resulting in inferior overall performance metrics.

Next, visual examples of CT reconstruction results at CTCA positions are presented for both methods. Figure 5.15 illustrates slice reconstructions using the Interpolation-Based method, depicted in the rightmost slice. Then, Figure 5.16 displays the reconstructed slice output from the CNN-Based method, specifically the tested U-Net model, in the rightmost slice. In all images, the leftmost image shows the lower CT input slice, while the middle image displays the upper CT input slice. In these examples, the reconstructed slice is closer to the lower input slice than to the

upper input slice.

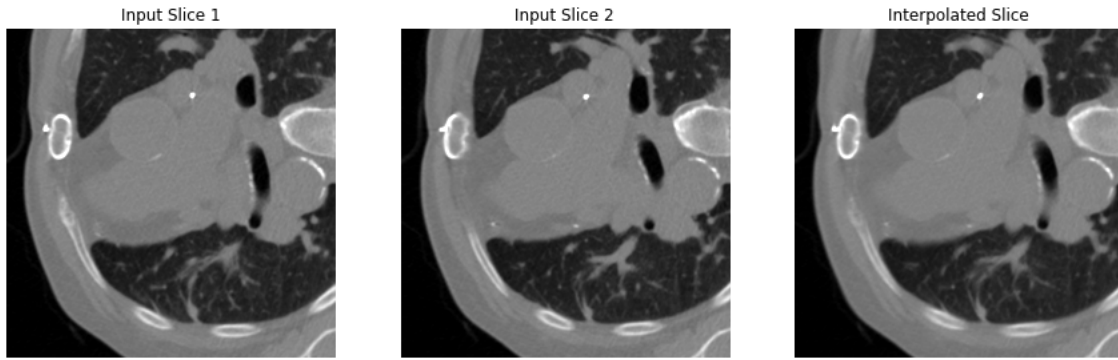


Figure 5.15: Result example of slice reconstruction for the Interpolation-Based method.

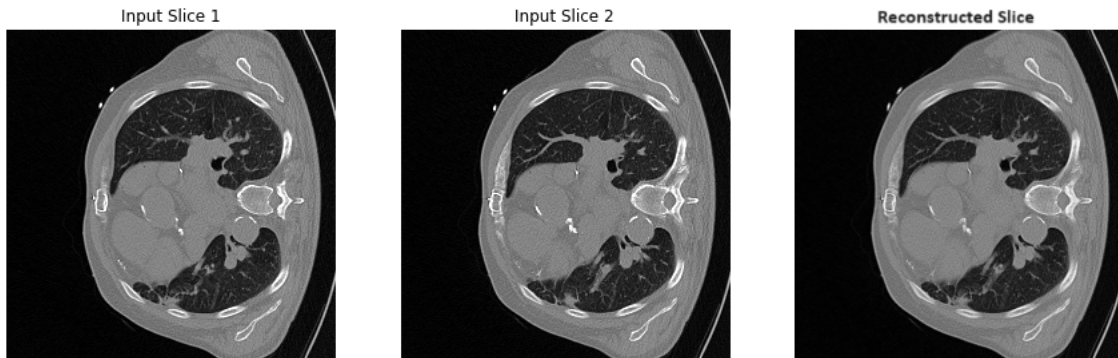


Figure 5.16: Result example of slice reconstruction for the CNN-Based method.

Given that the methods results in a reconstructed CT slice at a position where there is correspondence to a CTCA slice, the examples in Figures 5.17 and 5.18 presents the reconstructed CT slice along with the input CT slices and, on the right, the corresponding CTCA slice at the reconstruction position is shown.

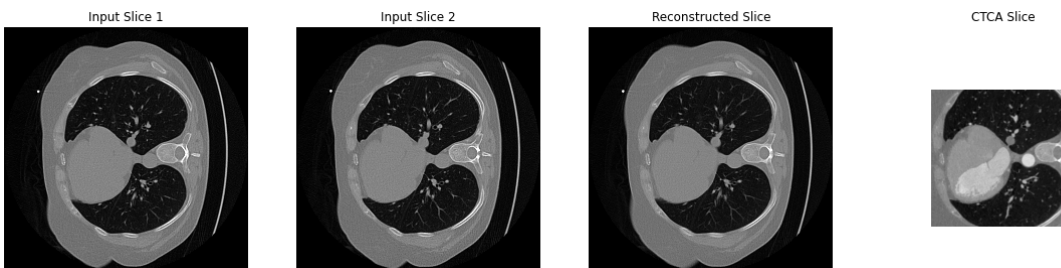


Figure 5.17: Result example of slice reconstruction for the Interpolation-Based method, along with the corresponding CTCA slice.

In Figure 5.19, side by side, the results of reconstructing CT slices at CTCA-derived positions using both the Interpolation-Based method and CNN-Based method are presented, allowing

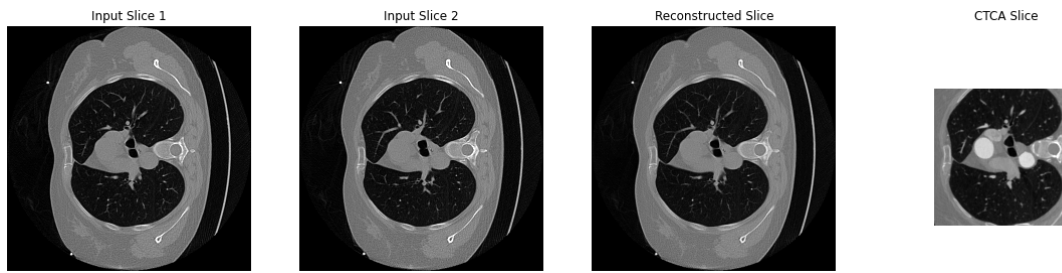


Figure 5.18: Result example of slice reconstruction for the CNN-Based method, along with the corresponding CTCA slice.

a visual comparison. This example illustrates the same slice reconstructed by both methods. The corresponding CTCA slice is the slice used for anatomical comparison, positioned at the same z-position where the CT slices were reconstructed by both methods.

Upon examining the examples, it is evident that both methods demonstrate a successful fusion of the two input slices, with the reconstructed slice bearing a stronger resemblance to the lower input slice due to its closer proximity. This is evident in the more pronounced characteristics of the lower slice within the reconstructed image. The reconstructed slice effectively bridges the gap between the adjacent input slices, featuring clear and well-defined internal structures.

The weighted averaging technique combines information from both input slices, emphasizing the contributions of the slice closer to the reconstruction position. Consequently, the final reconstructed slice appears smooth and coherent, capturing anatomical details that were previously unavailable in the original CT data.

This process enhances the axial resolution of the CT scans, resulting in high-quality reconstructions that maintain anatomical accuracy and provide clearer and more detailed images.

5.3.2 Autoencoder with CTCA information

In the previous methods, the information from the CTCA data was solely the z-position of the CTCA slice used to reconstruct the CT slice through the weighted averaging technique.

In this case, as explained earlier, the AE aims to incorporate CTCA information into the model input to complement the CT input data, thereby improving the CT reconstruction quality derived from previous models. Specifically, the inputs included the result from the CT reconstruction and the corresponding CTCA at that position. The goal was to integrate contextual information from the CTCA data to enhance the reconstruction accuracy of the CT slice.

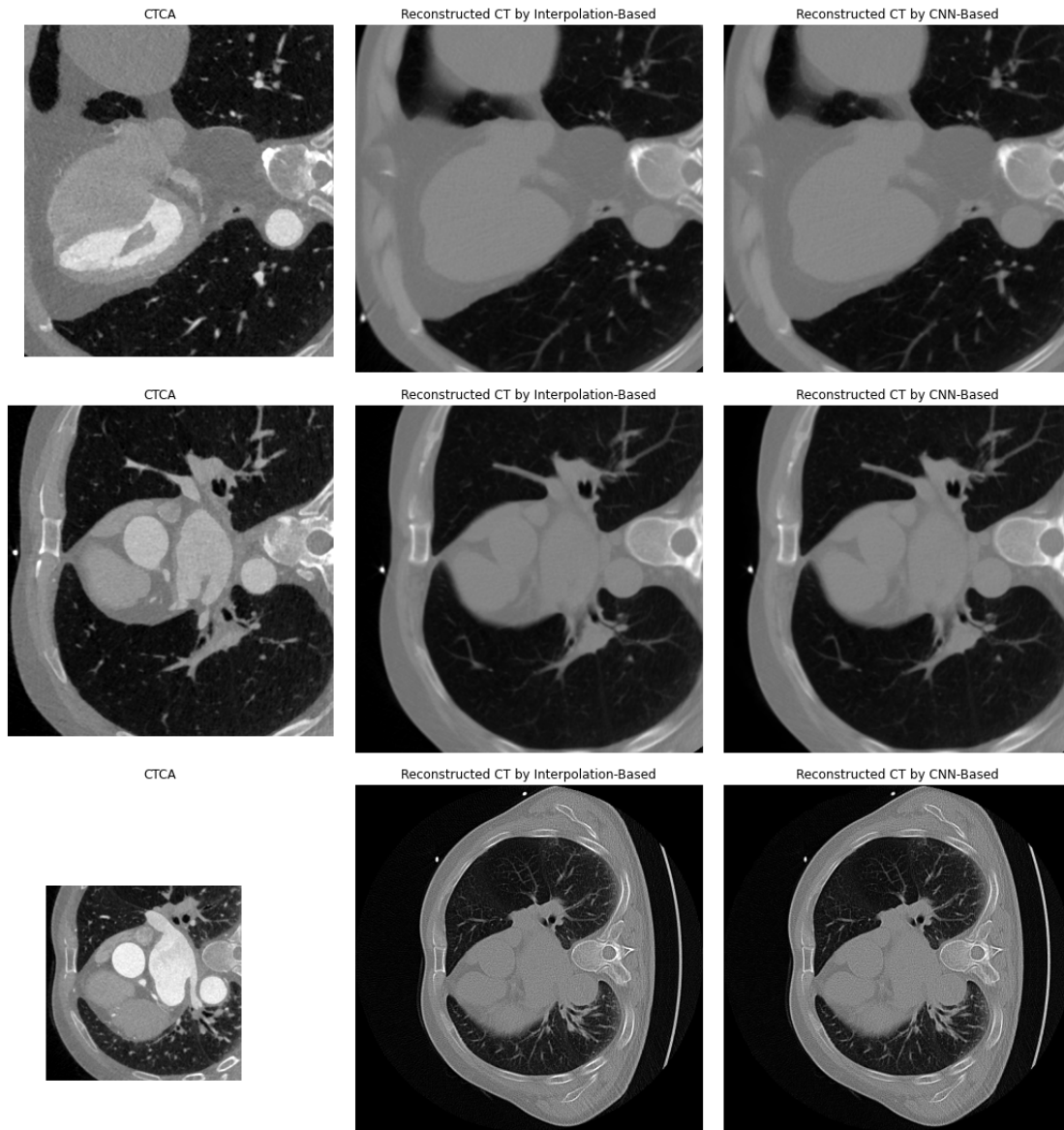


Figure 5.19: Results examples of slice reconstruction using the Interpolation-Based method (middle) and the CNN-Based method (right), along with the corresponding CTCA slice(left).

Table 5.2 presents the metric results obtained for the AEs tested with input data derived from the previous Interpolation-Based method and the CNN-Based method, along with the results from the previous methods.

In an overarching view of the metrics results, it can be affirmed that the AE tested as input using the result from the CNN-Based method of CT reconstruction demonstrates superior performance compared to the AE using the output from the interpolation-based method. This is evidenced by all evaluated metrics showing better values for this AE, including higher DSC score and lower FID and FID* scores. This outcome was expected as this method aimed to enhance the previous reconstruction, and in this case, the reconstruction obtained by the CNN-Based method was also the superior one.

Table 5.2: Results of the CT Reconstruction methods VS AE with CTCA information methods.

Method	DSC	FID	FID*
1. Interpolation- Based	0.709 ± 0.147	7.6222	7.6222
2. CNN- Based	0.719 ± 0.121	6.8898	6.8898
3. AE w/ CT Interpolation Input	0.713 ± 0.138	48.7646	32.9742
4. AE w/ CT CNN Input	0.743 ± 0.139	44.5872	30.6835

Comparing the results obtained for the AEs with the results from CT reconstruction at CTCA positions, as detailed in the preceding section, it is evident that quantitatively, exhibits superior performance in terms of the DSC compared to the CT reconstruction method. This implies that integrating CTCA information into the reconstruction process effectively enhances the quality of CT image reconstruction. However, concerning the FID values, there is a significant increase with the AE approach, indicating that the reconstructed images do not faithfully capture the statistical and structural characteristics of the real images. This discrepancy suggests a lower overall reconstruction quality compared to the method focused solely on CTCA-derived positions.

The distinction between the FID and FID* values emphasizes that integrating CTCA information into specific regions of the CT slice results in discernible differences between the reconstructed regions with and without CTCA data integration. Thus, the calculation of the FID* value specifically for the CT slice region corresponding to the CTCA proved useful. This adjustment was crucial because CTCA data were not available for the entire CT dimension, so the information was integrated only into a specific region of the CT slice. Consequently, this could result in differences between the regions with and without CTCA data integration in the output.

Notably, the FID* value calculated for the CT slice region corresponding to the CTCA is notably lower than the FID calculated for the entire CT slice, underscoring these disparities. However, even the FID* values are significantly higher for the AEs compared to the CT reconstruction methods at CTCA positions without CTCA data integration. Thus, while the integration of CTCA information appears beneficial based on DSC metric, the higher FID value suggests that this integration does not enhance the statistical similarity between the reconstructed images and the real images.

Visual examples of AE results further elucidate these findings. Figures 5.20 and 5.21 present examples of slice reconstruction outcomes using AE with CTCA information, utilizing CT inputs derived from the interpolation-based method and the CNN-based method, respectively. Each illustration includes the input slice (left), the reconstructed slice by AE (middle), and the input CTCA slice corresponding to the position where the input CT slice was reconstructed (right), providing visual context to complement the quantitative analysis.

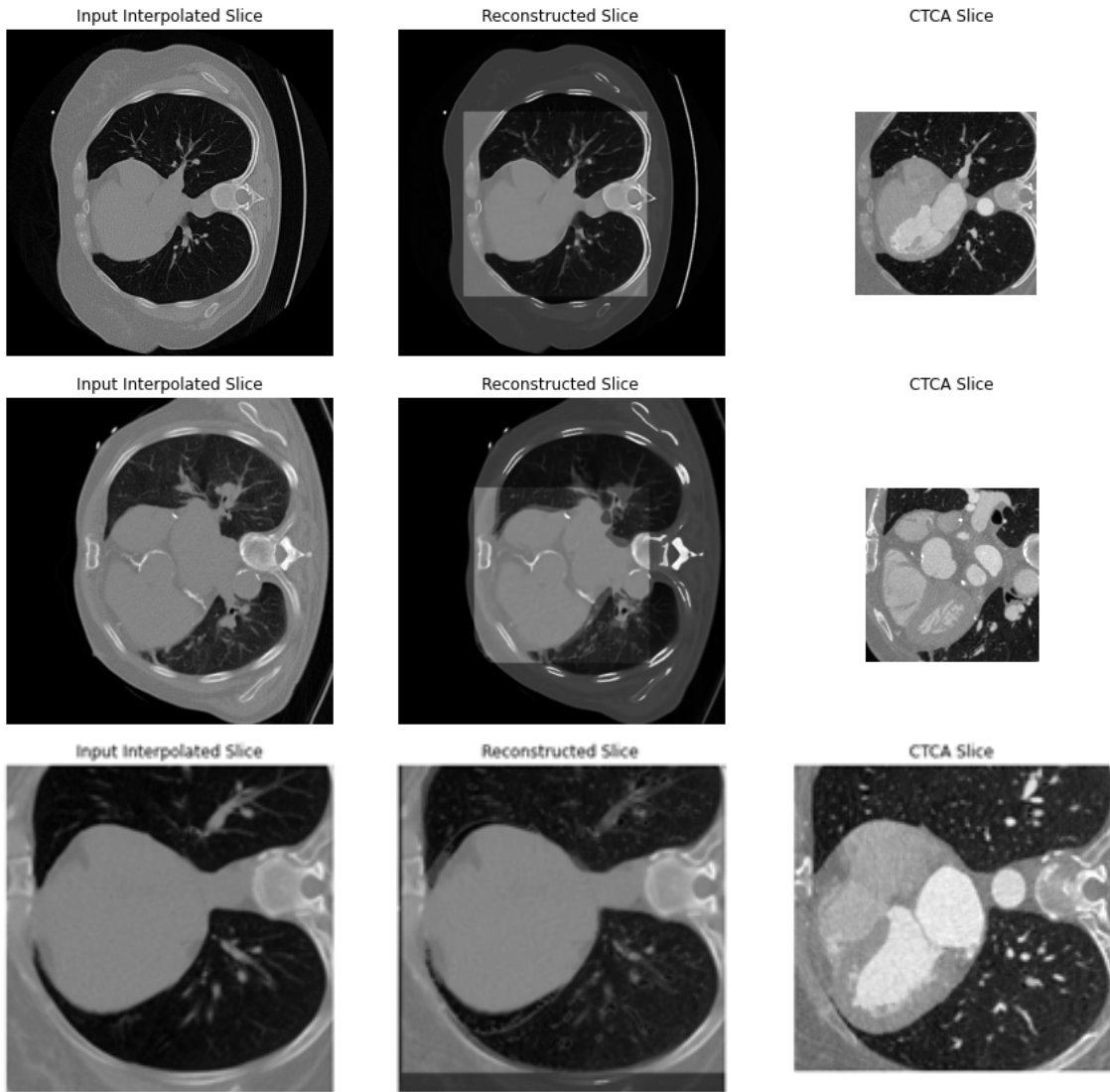


Figure 5.20: Results examples of slice reconstruction for AE with CTCA information, using as CT inputs the results from the interpolation-based method.

Upon close examination of the visual examples, the quantitative results become more comprehensible.

Primarily, as concluded before, there is a clear visual distinction that underscores the differences between the reconstructed regions with and without CTCA data integration. These disparities between within and outside of the integrated region are noticeable, which rationalizes the high FID value calculated for the entire CT slice. However, despite the FID* value for the CT slice region corresponding to the CTCA being notably lower than the overall FID value, it remains relatively high compared to previous methods. This highlights a limitation of the model associated with integrating CTCA information into the CT reconstruction process that should be taken into consideration when interpreting the FID and also the DSC results obtained.

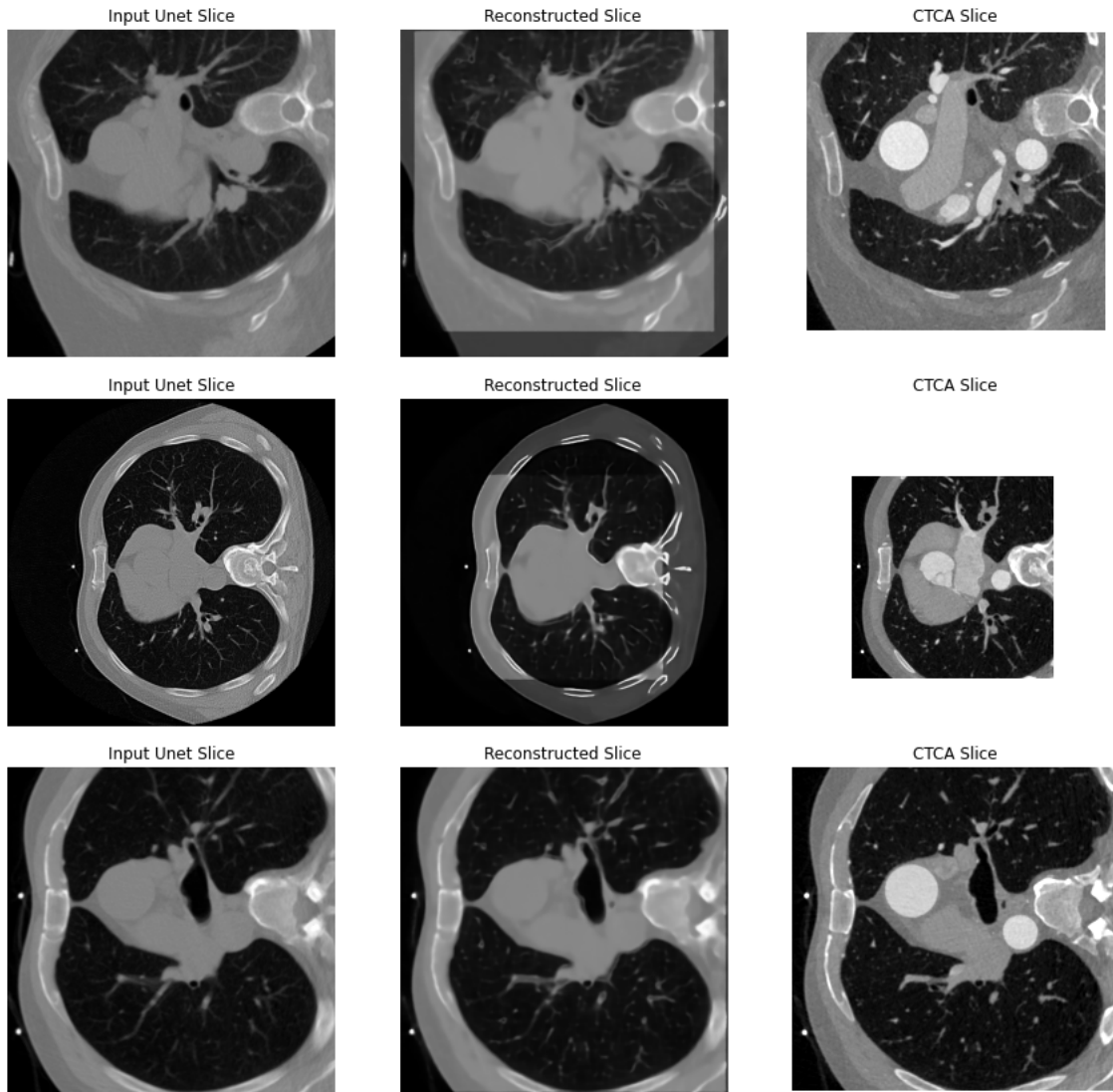


Figure 5.21: Results examples of slice reconstruction for AE with CTCA information, using as CT inputs the results from the CNN-based method.

Examining the contours of the reconstructed images, particularly the boundaries of the myocardium, it becomes evident that the limitation stems due to misalignment between the CTCA, at the position where the CT slice was reconstructed, and the resulting CT reconstruction by AEs incorporating CTCA information. Specifically, it is clear that integrating CTCA information is causing some anomalous effects in the more misaligned images.

Although the alignment module was applied to CTCA images in order to overcome differences in spatial resolution between CT scans and their respective CTCA counterparts across all axes, upon examining these cases, it becomes perceptible that there is still room for improvement in alignment. With the alignment strategy implemented, it is generally observed that the CTCA

slices are well aligned with the CT slices. However, it's important to acknowledge that the CT and CTCA images were not acquired simultaneously, and so the patient's respiratory motion during the exams is not synchronized between both scans. This variability results in minor misalignments, which may be more noticeable in certain patients than in others. Therefore, since the AE network is a straightforward model that takes both CT and CTCA inputs, computes their feature representations, and blindly sums them for CT reconstruction without learning to handle potential misalignments between the images, it is expected that when the CTCA slice is perfectly aligned with the CT, the AE's reconstruction of CT works well, enhancing the input reconstruction. However, in cases where even minor misalignments occur, as observed in the examples shown, it leads to less satisfactory results.

Therefore, the elevated FID* values observed are understandable, as the CT reconstructions by the AEs exhibit pixels in unexpected positions, thereby reducing the realism and quality of the reconstruction and consequently increasing the FID score.

On the other hand, while the visual inspection of examples clearly shows limitations in the CT reconstructions by the AEs, resulting in poorer reconstruction quality, quantitatively, particularly in terms of the DSC metric, it was found that the results were better compared to the CT Reconstruction methods, which is contradictory. The explanation for this discrepancy lies in the training process of the AEs, where the DSC metric was somewhat optimized during the training. This is because the loss function calculation was performed between the input CTCA and the network output, specifically for HU values below 0. Then, when computing the DSC, a threshold was applied below 0 HU, and the overlap of the resulting masks was calculated. Therefore, fat pixels in corresponding regions of the CTCA, appearing in the network output (i.e., in the reconstructed slice by the AE), contribute to a higher overlap.

This dual observation underscores the nuanced evaluation needed when assessing the performance of AEs integrating CTCA information for CT reconstruction. While quantitative metrics like DSC may suggest improved performance, the presence of artifacts and misalignments in the reconstructed CT images, reflected in higher FID* values, highlights the complexity and ongoing challenges in achieving high-fidelity CT reconstructions with integrated CTCA data.

5.3.3 Connected Auto-Encoders

The application of the CAEs focuses on integrating information from CTCA slices into the reconstruction of CT slices at corresponding CTCA positions. This approach implements principles of image separation to isolate the contrast from the CTCA, thereby aiming to achieve an improved reconstruction of the CT images. Initially, this methodology was implemented in a 2D scenario. Subsequently, the approach was extended from a 2D to a volumetric framework, with the aim of reconstructing CT images along with contrast and CTCA volumes. This extension uses two

CT slices as input along with all intervening CTCA slices, with the output expected to match the number of slices in the input CTCA volume.

Table 5.3 summarizes the quantitative results obtained for the metrics calculated for all the methods implemented and tested in this chapter.

Table 5.3: Summarization of the results of the different methods.

Method	DSC	FID	FID*
1. Interpolation- Based	0.709 ± 0.147	7.6222	7.6222
2. CNN- Based	0.719 ± 0.121	6.8898	6.8898
3. AE w/ CT Interpolation Input	0.713 ± 0.138	48.7646	32.9742
4. AE w/ CT CNN Input	0.743 ± 0.139	44.5872	30.6835
5. 2D CAEs	0.732 ± 0.126	51.2942	25.8851
6. 3D CAEs	0.749 ± 0.125	84.4014	55.0369

The 2D CAEs method achieved a DSC of 0.732 ± 0.126 indicating significant overlap between the reconstructed CT images and their respective CTCA masks. Compared to previously evaluated methods, this AE demonstrates an improvement in DSC over the CNN-Based CT Reconstruction approach. This improvement suggests successful enhancement in the specific task of CT image reconstruction. However, the DSC value is slightly lower compared to the AE incorporating CTCA information, but this difference may not be statistically significant, especially considering the higher DSC observed in the latter method, which could be attributed to the limitations previously discussed.

The FID score of 31.2942 suggests that the generated images are closer to real images compared to using the AE with CTCA information, but there are still noticeable differences. In this case, the difference between FID and FID* is less pronounced, indicating that the differences between reconstructed regions with and without CTCA data integration are more subtle. This may be attributed to the increased complexity of this AE compared to previous methods, which could have facilitated slight improvements. However, it is essential to note that the limitations mentioned earlier also influenced this approach, hence the relatively high FID values.

In its turn, the 3D CAEs achieved a DSC of 0.749 ± 0.125 , representing a slight improvement over the 2D approach, which obtained a DSC of 0.732 ± 0.126 . This marginal increase suggests that transitioning to a 3D framework has the potential to enhance the preservation of anatomical structures and improve the accuracy of volumetric reconstruction.

However, the FID for the 3D approach was notably higher at 84.4014, compared to 31.2942 for the 2D approach. This increase suggests that the images generated by the 3D model are further from real images in terms of feature space distance, highlighting challenges in accurately representing spatial and textural details in 3D reconstructions. The higher FID* value for the 3D

approach further emphasizes the difficulty in integrating CTCA information into specific regions of the CT slice without introducing discernible differences between reconstructed regions with and without CTCA data integration.

Visual results provide further insights into the performance of the models. Figure 5.22 depicts input CT and CTCA slices alongside their respective reconstructions and the extracted contrast information, for the 2D AE. In the examples provided, the input slices to the network are depicted above: the CT slice reconstructed at a CTCA position using the U-Net with weighted average in the feature representations (CNN-Based CT Reconstruction), along with its corresponding CTCA slice. Below are the outputs, namely the reconstruction of the input CT slice, the reconstruction of the input CTCA slice, and the slice representing the contrast present in the CTCA.

Regarding the reconstruction of CT and CTCA, upon comparing its input slices with output slices, it is evident that the model successfully reconstructs the images with a high degree of fidelity. The anatomical structures are accurately preserved, indicating that the encoder-decoder network effectively captures and reconstructs the essential features of the images.

However, the reconstruction of the CT images appears to result in some loss of fine details and presents lighter tones, which seems to counterbalance the darker tones observed in the contrast reconstruction. This adjustment is likely aimed at ensuring that the combined reconstruction from these two outputs closely resembles the reconstruction of the CTCA images.

Furthermore, while the contrast reconstruction output demonstrates the model's ability to highlight and reconstruct contrast information separately from the CT data, it does not effectively isolate the contrast present in the CTCA. This suggests a need for improvement in the model's ability to specifically extract and represent contrast elements that are unique to CTCA images.

In Figures 5.23 and 5.24, the results of the various reconstructions obtained by the 3D AE are presented: CT' represents the reconstructed N slices of CT, CTCA' shows the reconstruction of the N slices of CTCA, and CONT denotes the corresponding contrast reconstruction. Additionally, it includes the input slices to the model from CT and CTCA.

Upon analyzing the images, it is evident that the 3D CAEs approach has effectively integrated CTCA data to enhance CT reconstruction at CTCA positions, distinguishing between contrast and CT reconstructions. However, there are instances where reconstructed CT slices exhibit broader ranges, in order to balance the loss between combined contrast and CTCA reconstruction and the CTCA reconstruction against the loss between the original CT input and the sampled CT reconstruction for 2 slices, resulting in higher FID values.

The elevated FID values indicate that the model's generated images deviate further from real images in terms of feature space distance. This challenge likely stems from difficulties in accurately representing spatial and textural details in 3D reconstructions, particularly when integrating contrast information while maintaining fidelity to the original CT image. Despite these challenges,

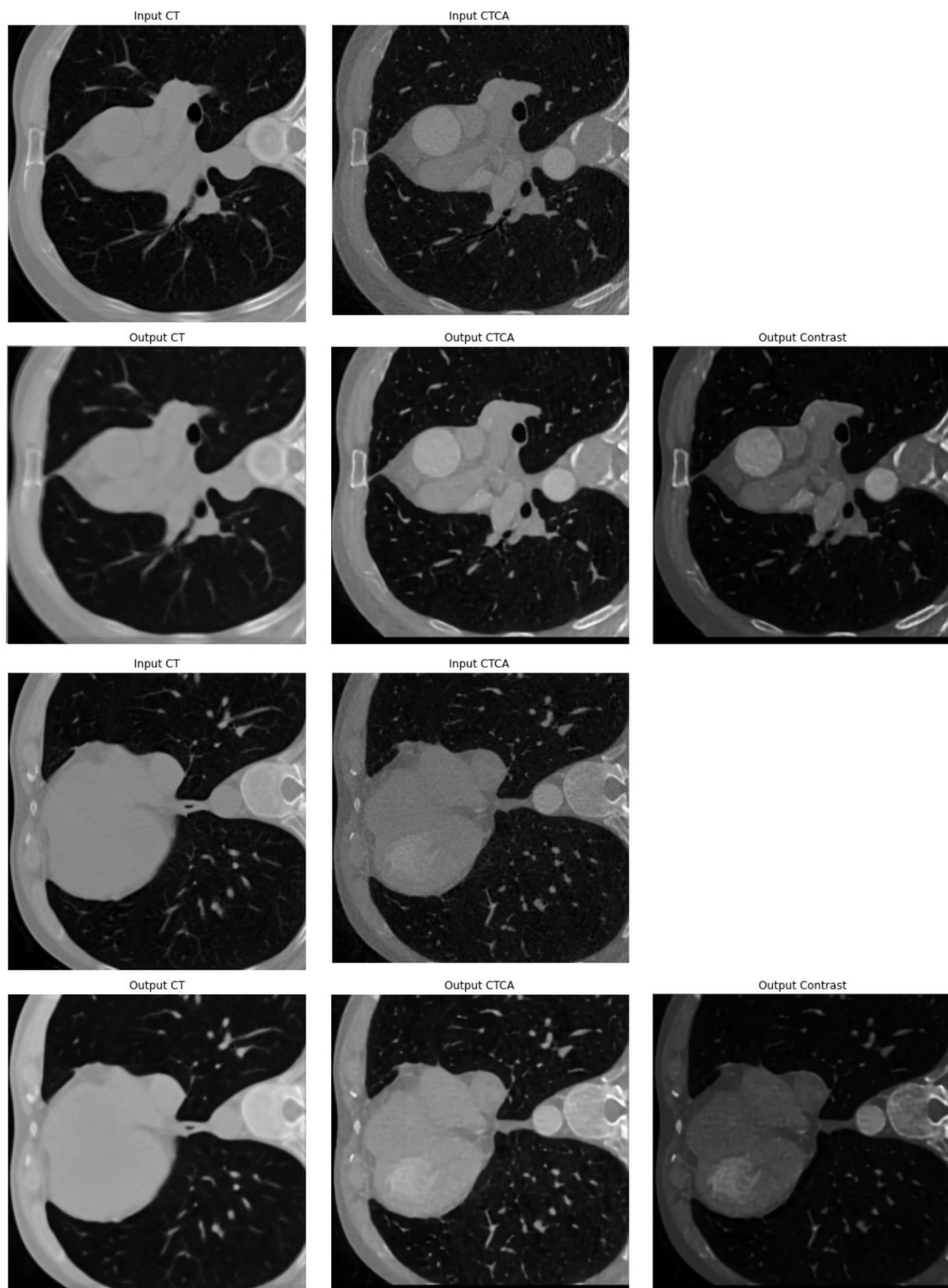


Figure 5.22: Results examples of slice reconstruction for the 2D CAEs.

the high DSC achieved by the model suggests substantial overlap between reconstructed CT images and their respective CTCA masks. In this regard, the model appears to have learned better to avoid reconstructing pixels in CT that correspond to areas from CTCA but do not align with the

original CT image, marking an improvement over the AE model with CTCA integration, where such discrepancies may have been more prevalent.

This enhancement suggests that the 3D CAEs approach effectively utilizes CTCA data to inform the reconstruction process, ensuring that the generated CT images closely adhere to the original anatomical structures depicted in the CT scans. By integrating CTCA information into CT reconstruction and employing principles of image separation, this approach enhances the accuracy and utility of reconstructions by treating volumetric data as a cohesive entity rather than isolated slices.

Nevertheless, visible constraints and limitations indicate that future refinements would be beneficial to fully exploit the potential of this method.

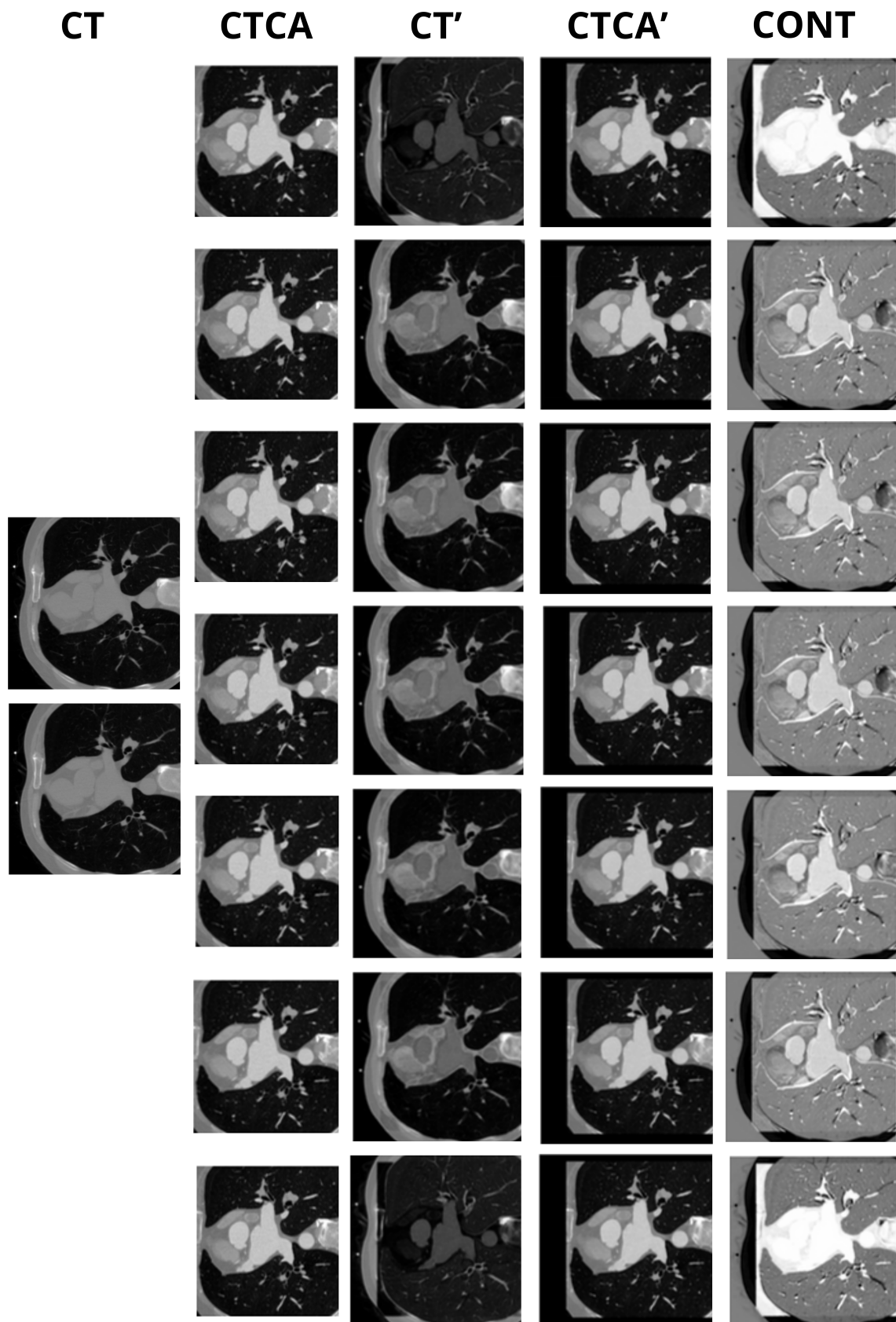


Figure 5.23: Result example of slice reconstruction for the 3D CAEs. The first column showcases the two input CT slices, followed by the second column depicting the N input CTCA slices located between the CT input slices. The third column displays the reconstructed CT images, the fourth column shows the reconstructed CTCA images, and the final column illustrates the reconstructed contrast. Each line corresponds to slices from 1 to N .

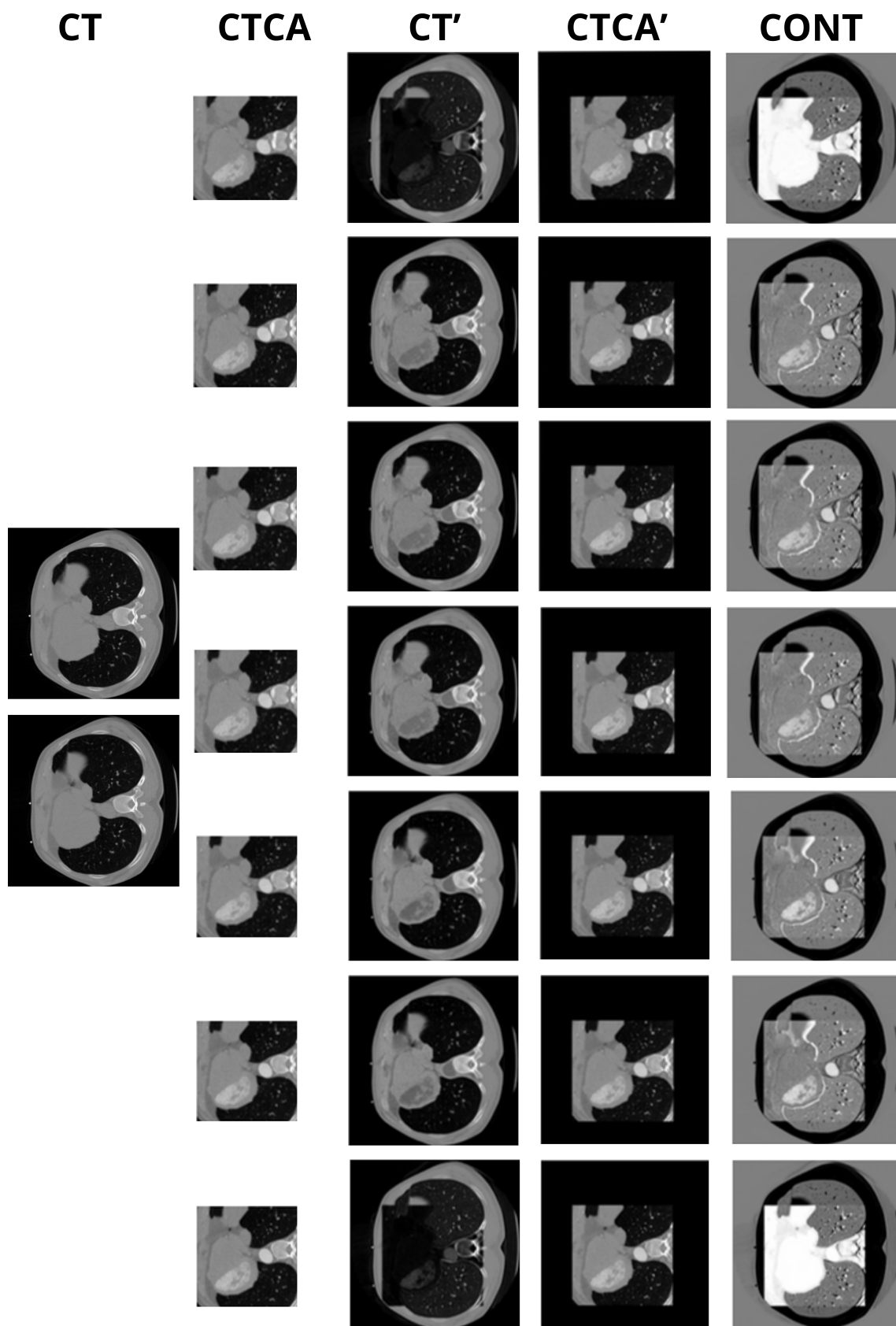


Figure 5.24: Result example of slice reconstruction for the 3D CAEs. The first row showcases the two input CT slices, followed by the second row depicting the N input CTCA slices located between the CT input slices. The third row displays the reconstructed CT images, the fourth row shows the reconstructed CTCA images, and the final row illustrates the reconstructed contrast. Each line corresponds to slices from 1 to N.

Chapter 6

Conclusions and Future Work

Combining analysis of CT and CTCA scans using deep learning approaches has the potential to provide additional insights. However, this potential is hindered by the varying imaging resolutions of CT and CTCA scans, as well as slice misalignments due to patient motion during the exams. This dissertation focused on addressing these challenges through the development of robust and automated tools for the alignment and super-resolution of CT and CTCA scans. The motivation is to enhance the diagnostic value of CT scans, which are currently underutilized in routine clinical practice.

This investigation towards SR reconstruction in cardiac CT imaging explored two main strategies. The first approach prioritized leveraging solely CT data to reconstruct missing slices within the known locations of the existing CT data. This allowed direct validation of the SR reconstruction methods, assessing their accuracy and feasibility by comparing them to the original slices in the full-resolution CT data. It highlighted the potential of CNNs-based models specifically for CT slice reconstruction, as the U-Net model clearly outperformed the interpolation method, achieving superior detail preservation and overall reconstruction quality. This success sets a solid foundation for further exploration of U-Net’s application in real-world cardiac CT imaging scenarios.

Building upon the success of the first strategy, the second approach integrates CTCA data. This crucial step allows for the generation of super-resolved CT reconstructions, including previously unavailable slices at specific z-positions at corresponding CTCA positions. This approach successfully addresses the overarching motivation of this dissertation by effectively achieving high-resolution CT scans that match the resolution level of CTCA data. When reconstructing slices corresponding to CTCA positions based on distances from the desired positions of the generated slice and the input slices, it becomes evident that utilizing a U-Net model yields superior results, akin to the previous approach adaptation. This approach significantly enhances CT scan axial resolution, yielding high-quality reconstructions with exceptional detail. Notably, the realism achieved with these methods matches that of the previous approach, demonstrating consistent performance even for reconstructing slices originally absent from the dataset. This underscores the robustness and applicability of these techniques, emphasizing their potential for broader applications in CT imaging.

However, when exploring the integration of CTCA information into AEs and CAEs to enhance CT reconstruction quality, it's evident that the remaining methods exhibit increased complexity. While this integration effectively improves the DSC evaluation metric due to the incorporation of CTCA information, it comes at the expense of significantly diminished realism. This issue stems from alignment limitations, which become more pronounced in certain patients.

The developed alignment module between CT and CTCA images handles alignment across all axis, including truncating the CT FOV to match that of CTCA, z-axis alignment, and aligning CTCA in the xy plane based on CT coordinates. However, due to the non-simultaneous acquisition of CT and CTCA scans, inherent variations in patient breath-holding can introduce residual misalignment. This highlights the need for further refinement of the alignment step, crucial for optimal performance of CTCA-integrated methods.

The application of these methods envisioned that a sufficiently large network could semantically interpret both CT and CTCA images, intelligently fusing the information rather than blindly summing it, eliminating these inconsistencies through alignment and summation. However, the straightforward implementation of these methods merely extracts features from each image and blindly sums them up, without learning to handle potential misalignments between the images and leading to the observed inconsistencies in CT reconstruction.

To overcome the alignment limitations, a future improvement proposal would involve developing a deformable alignment module first and then focusing on the SR task, rather than attempting to perform both steps in an end-to-end learning approach. Currently, only a rigid alignment module is employed, which is insufficient due to the inherent structural variations within patient images. These variations extend beyond spatial positions and necessitate the incorporation of dynamic alignment. This dynamic alignment would register the CT image to achieve precise correspondence with the CTCA image. Once this dynamic CT-CTCA alignment is achieved, the developed methods could be applied without encountering the observed limitations.

The introduction of 3D CAEs for handling volumetric data marks a significant advancement. Notably, the 3D CAE approach effectively integrates CTCA data while adhering to principles of image separation. This demonstrates its ability to treat volumetric data as a whole, paving the way for further exploration of 3D frameworks for achieving super-resolved cardiac CT reconstructions.

In conclusion, the research presented in this dissertation has successfully addressed its objectives, advancing the state-of-the-art in cardiac CT imaging and CAD diagnosis. By enhancing the axial resolution and alignment of CT and CTCA scans, the proposed methods have the potential to significantly improve diagnostic accuracy and efficiency, ultimately benefiting clinical practice and patient outcomes. Future research directions also include investigating the use of GANs and diffusion models for artificial CTCA generation from NCCT data, a poorly explored field since existing approaches focus on individual 2D slice generation, lacking the integration of 3D information.

References

- [1] JP Duggan, AS Peters, GD Trachiotis, and JL Antevil. Epidemiology of coronary artery disease. *Surg Clin North Am*, 102(3):499–516, 2022.
- [2] Raja D. Shahjehan and Bilal S. Bhutta. Coronary artery disease. In *StatPearls*. 2023.
- [3] U. J. Schoepf, C. R. Becker, B. M. Ohnesorge, and E. K. Yucel. CT of coronary artery disease. *Radiology*, 232(1):18–37, 2004.
- [4] European Society of Cardiology. Guidelines & scientific documents. coronary heart disease, 2023.
- [5] Luca Pugliese, Francesca Ricci, Giacomo Sica, Mariano Scaglione, and Salvatore Masala. Non-contrast and contrast-enhanced cardiac computed tomography imaging in the diagnostic and prognostic evaluation of coronary artery disease. *Diagnostics*, 13(12), 2023.
- [6] Shee Yen Tay, Po-Yen Chang, Wilson T Lao, Ying Chin Lin, Yi-Han Chung, and Wing P Chan. The proper use of coronary calcium score and coronary computed tomography angiography for screening asymptomatic patients with cardiovascular risk factors. *Scientific reports*, 7(1):17653, 2017.
- [7] Sungjin Lee, Beanbonyka Rim, Sung-Shick Jou, Hyo-Wook Gil, Xibin Jia, Ahyoung Lee, and Min Hong. Deep-learning-based coronary artery calcium detection from CT image. *Sensors*, 21(21):7059, 2021.
- [8] W. K. Cheung, R. Bell, A. Nair, L. J. Menezes, R. Patel, S. Wan, K. Chou, J. Chen, R. Torii, R. H. Davies, J. C. Moon, D. C. Alexander, and J. Jacob. A computationally efficient approach to segmentation of the aorta and coronary arteries using deep learning. *IEEE Access: Practical Innovations, Open Solutions*, 9:108873–108888, 2021.
- [9] Adam Saxton, Rameez Chaudhry, and Bobby Manna. Anatomy, thorax, heart right coronary arteries. In *StatPearls*. 2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK537357/>.
- [10] Myron Prinzmetal, Eliot Corday, H. C. Bergman, Lois Schwartz, and Ramon J. Spritzler. Radiocardiography: A new method for studying the blood flow through the chambers of the heart in human beings. *Science*, 108(2804):340–341, 1948.
- [11] Pediatric Heart Specialists. Anatomy of the human heart. Online, 2023.
- [12] Newport Cardiothoracic and Transplant Surgery. Coronary artery surgery. Online, 2023.
- [13] Circulatory system. Online, 2023.

- [14] Rai Dilawar Shahjehan and Beenish S Bhutta. Coronary artery disease. *StatPearls*, 2023.
- [15] Lacramioara Ionela Butnariu, Laura Florea, Minerva Codruta Badescu, Elena Țarca, Irina-Iuliana Costache, and Eusebiu Vlad Gorduza. Etiologic puzzle of coronary artery disease: How important is genetic component? *Life*, 12(6):865, 2022.
- [16] Jacob Fog Bentzon, Fumiyuki Otsuka, Renu Virmani, and Erling Falk. Mechanisms of plaque formation and rupture. *Circulation Research*, 114(12):1852–1866, 2014.
- [17] R. B. Singh, S. A. Mengi, Y. J. Xu, A. S. Arneja, and N. S. Dhalla. Pathogenesis of atherosclerosis: A multifactorial process. *Exp Clin Cardiol*, 7(1):40–53, 2002. PMID: 19644578; PMCID: PMC2716189.
- [18] The Patient Guide to Heart. Coronary artery disease. Online, 2023.
- [19] Udaya Ralapanawa and Ramiah Sivakanesan. Epidemiology and the magnitude of coronary artery disease and acute coronary syndrome: A narrative review. *Journal of Epidemiology and Global Health*, 11(2):169–177, 2021.
- [20] James E. Dalen, Joseph S. Alpert, Robert J. Goldberg, and Ronald S. Weinstein. The epidemic of the 20th century: Coronary heart disease. *The American Journal of Medicine*, 127(9):807–812, 2014.
- [21] European Heart Network. European cardiovascular disease statistics. Online, 2023.
- [22] É. Puymirat. Épidémiologie de la maladie coronarienne [Epidemiology of coronary artery disease]. *La Revue du praticien*, 65(3):317–320, 2015.
- [23] John C. Brown, Thomas E. Gerhardt, and Emily Kwon. Risk factors for coronary artery disease. In *StatPearls*. StatPearls Publishing, 2023.
- [24] Andrea C. Skelly, Rebecca Hashimoto, Daniel I. Buckley, and et al. *Noninvasive Testing for Coronary Artery Disease*. Number 171 in Comparative Effectiveness Reviews. Agency for Healthcare Research and Quality (US), Rockville, MD, 2016.
- [25] G. D. Rubin. Emerging and evolving roles for CT in screening for coronary heart disease. *Journal of the American College of Radiology : JACR*, 10(12):943–948, 2013.
- [26] D. K. Arnett, R. S. Blumenthal, M. A. Albert, A. B. Buroker, Z. D. Goldberger, E. J. Hahn, C. D. Himmelfarb, A. Khera, D. Lloyd-Jones, and McEvoy. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: A report of the american college of cardiology/american heart association task force on clinical practice guidelines. *Circulation*, 140(11):596–646, 2019.
- [27] Alexander Chua, Ron Blankstein, and Brian Ko. Coronary artery calcium in primary prevention. *Australian Journal for General Practitioners*, 49(8):464–469, 2020.
- [28] Ahmed C. Razavi, Arthur S. Agatston, Leslee J. Shaw, Carlo Nicola De Cecco, Maroeska van Assen, Laurence S. Sperling, Marcio S. Bittencourt, Melissa A. Daubert, Khurram Nasir, Roger S. Blumenthal, Martin B. Mortensen, Seamus P. Whelton, Michael J. Blaha, and Omar Dzaye. Evolving role of calcium density in coronary artery calcium scoring and atherosclerotic cardiovascular disease risk. *JACC. Cardiovascular Imaging*, 15(9):1648–1662, 2022.

- [29] Ying Wang, Walter R. Witschey, and David A. Bluemke. Imaging cardiovascular calcification. *Journal of the American Heart Association*, 7(13):e008564, Jun 2018.
- [30] Simon Deseive, Francesca Pugliese, Aloha Meave, Erick Alexánderson, Stefan Martinoff, Martin Hadamitzky, Steffen Massberg, and Jörg Hausleiter. Image quality and radiation dose of a prospectively ECG-triggered high-pitch data acquisition strategy for coronary CT angiography: The multicenter, randomized protection IV study. *Journal of Cardiovascular Computed Tomography*, 9, 2015.
- [31] Ricardo C. Cury, Jonathon Leipsic, Suhny Abbara, Stephan Achenbach, Daniel Berman, Marcio Bittencourt, Matthew Budoff, Kavitha Chinnaiyan, Andrew D. Choi, Brian Ghoshhajra, Jill Jacobs, Lynne Kowek, John Lesser, Christopher Maroules, Geoffrey D. Rubin, Frank J. Rybicki, Leslee J. Shaw, Michelle C. Williams, Eric Williamson, Charles S. White, and Ron Blankstein. CAD-RADS 2.0 - 2022 coronary artery disease-reporting and data system: An expert consensus document of the society of cardiovascular computed tomography (SCCT), the american college of cardiology (ACC), the american college of radiology (ACR), and the north america society of cardiovascular imaging (NASCI). *Journal of Cardiovascular Computed Tomography*, 16(6):536–557, 2022.
- [32] Robert A. Burch, Tariq A. Siddiqui, Li Cheng Tou, Katherine B. Turner, and Muhammad Umair. The cost effectiveness of coronary CT angiography and the effective utilization of CT-fractional flow reserve in the diagnosis of coronary artery disease. *J Cardiovasc Dev Dis*, 10(1):25, 2023.
- [33] Sorin Giusca, Marcel Schütz, Florian Kronbach, Dominik Wolf, Pascal Nunninger, and Grigorios Korosoglou. Coronary computer tomography angiography in 2021-acquisition protocols, tips and tricks and heading beyond the possible. *Diagnostics (Basel)*, 11(6):1072, 2021.
- [34] Jonathan R. Weir-McCall, Michelle C. Williams, Anoop S.V. Shah, Giles Roditi, James H.F. Rudd, David E. Newby, and Edward D. Nicol. National trends in coronary artery disease imaging: Associations with health care outcomes and costs. *JACC: Cardiovascular Imaging*, 16(5):659–671, 2023.
- [35] Abdullah Alomary, Abdelmoneim Sulieman, Mohammed Alsufayan, Fahad Alabdurazaq, Nada Faisal, Ameera Qari, Bandar M. Alanazi, Mohammed Alsaadi, Nissren Tamam, Essam Alkhybari, and David A. Bradley. Evaluation of radiation exposure for patients undergoing computed tomography perfusion procedure for acute ischemic stroke. *Radiation Physics and Chemistry*, 201:110447, 2022.
- [36] Ricky D. Turgeon and Tara Sedlak. Use of preventive medications in patients with nonobstructive coronary artery disease: Analysis of the promise trial. *CJC Open*, 3(2):159–166, 2020.
- [37] The Royal College of Radiologists. *Standards of Practice of Computed Tomography Coronary Angiography CTCA in Adult Patients*. British Society of Cardiovascular Imaging BSCI, UK, 2014.
- [38] K. Suh Young Kim. Prognostic value of coronary artery calcium scores from 1.5 mm slice reconstructions of electrocardiogram-gated computed tomography scans in asymptomatic individuals. *Scientific Reports*, 2022.
- [39] Fernando Uliana Kay. Dual-energy CT and coronary imaging. *Cardiovascular Diagnosis and Therapy*, 10(4):1090–1107, 2020.

- [40] S. Bruns, J.M. Wolterink, R.A.P. Takx, R.W. van Hamersvelt, D. Suchá, M.A. Viergever, T. Leiner, and I. Išgum. Deep learning from dual-energy information for whole-heart segmentation in dual-energy and single-energy non-contrast-enhanced cardiac CT. *Medical Physics*, 47(10):5048–5060, 2020.
- [41] Reza Forghani, Bruno De Man, and Rajiv Gupta. Dual-energy computed tomography: Physical principles, approaches to scanning, usage, and implementation: Part 1. *Neuroimaging Clinics of North America*, 27(3):371–384, 2017. Dual Energy CT: Applications in Head and Neck and Neurologic Imaging.
- [42] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE trans acoust speech signal process. Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29:1153 – 1160, 1982.
- [43] Xin Li and Michael Orchard. New edge-directed interpolation. *Image Processing, IEEE Transactions on*, 10:1521 – 1527, 2001.
- [44] Hui Ji and Cornelia Fermüller. Wavelet-based super-resolution reconstruction: Theory and algorithm. In *European Conference on Computer Vision (ECCV)*, volume 3954, 2006.
- [45] Yu-Wing Tai, Shuaicheng Liu, Michael S. Brown, and Stephen Lin. Super resolution using edge prior and single image detail synthesis. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2400–2407, 2010.
- [46] Noor Zaim Fariz Nor Azam, Haniza Yazid, and Saufiah Abdul Rahim. Super resolution with interpolation-based method: A review. *International Journal of Recent Advances in Multidisciplinary Research*, 9(2), 2022.
- [47] Pourya Shamsolmoali, M. Emre Celebi, and Ruili Wang. Deep learning approaches for real-time image super-resolution. *Neural Computing and Applications*, 32, 2020.
- [48] Jian Lu, Weidong Hu, and Yi Sun. A deep learning method for image super-resolution based on geometric similarity. *Signal Processing: Image Communication*, 70:210–219, 2019.
- [49] Kamal Nasrollahi and Thomas Moeslund. Super-resolution: A comprehensive survey. *Machine Vision and Applications*, 25:1423–1468, 2014.
- [50] F. Bruno Sixou. A review of the deep learning methods for medical images super resolution problems. *Innovation and Research in BioMedical Engineering*, 42(2):120–133, 2021.
- [51] Zhuoqun Yuan, Di Yang, Hongming Pan, and Yanmei Liang. Axial super-resolution study for optical coherence tomography images via deep learning. *IEEE Access*, 8:204941–204950, 2020.
- [52] Yukai Wang, Qizhi Teng, Xiaohai He, Junxi Feng, and Tingrong Zhang. CT-image of rock samples super resolution using 3D convolutional neural network. *Computers & Geosciences*, 133:104314, 2019.
- [53] Viren Jain. Adversarial image alignment and interpolation. *arXiv*, abs/1707.00067, 2017.
- [54] Junyoung Park, Donghwi Hwang, Kyeong Yun Kim, Seung Kwan Kang, Yu Kyeong Kim, and Jae Sung Lee. Computed tomography super-resolution using deep convolutional neural network. *Physics in Medicine and Biology*, 63(14):145011, 2018.

- [55] Cheng Peng, Wei-An Lin, Haofu Liao, Rama Chellappa, and S. Kevin Zhou. Deep slice interpolation via marginal super-resolution, fusion and refinement, 2019.
- [56] Akira Kudo, Yoshiro Kitamura, Yuanzhong Li, Satoshi Iizuka, and Edgar Simo-Serra. Virtual thin slice: 3D conditional GAN-based super-resolution for CT slice interval, 2019.
- [57] Cheng Peng, Wei-An Lin, Haofu Liao, Rama Chellappa, and S. Kevin Zhou. SAINT: spatially aware interpolation network for medical slice synthesis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 2020*, pages 7747–7756. IEEE, 2020.
- [58] Jiawei Li, Jae Chul Koh, and Won-Sook Lee. HRINET: Alternative supervision network for high-resolution CT image interpolation. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1916–1920, 2020.
- [59] Heng-Sheng Chao, Yu-Hong Wu, Linda Siana, and Yuh-Min Chen. Generating high-resolution CT slices from two image series using deep-learning-based resolution enhancement methods. *Diagnostics*, 12(11), 2022.
- [60] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014.
- [61] Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. GANs for medical image synthesis: An empirical study. *Journal of Imaging*, 9(3), 2023.
- [62] J. Frost. Mean squared error (MSE). <https://statisticsbyjim.com/regression/mean-squared-error-mse/>.
- [63] NRMSE: Normalized root mean square error. <https://rdrr.io/cran/INDperform/man/nrmse.html>.
- [64] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008.
- [65] Wikipedia. Structural similarity. https://en.wikipedia.org/wiki/Structural_similarity.
- [66] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402 Vol.2, 2003.
- [67] OECD. Catalogue of Tools & Metrics for Trustworthy AI, 2024.
- [68] Jason Brownlee. How to implement the inception score (IS) for evaluating GANs. *Generative Adversarial Networks, Machine Learning Mastery*, 2019. URL: [<https://machinelearningmastery.com/how-to-implement-the-inception-score-from-scratch-for-evaluating-generated-images/>].
- [69] Jason Brownlee. How to implement the frechet inception distance (FID) for evaluating GANs. *Generative Adversarial Networks, Machine Learning Mastery*, 2019. URL: [<https://machinelearningmastery.com/how-to-implement-the-frechet-inception-distance-fid-from-scratch/>].

- [70] Steve Göring, Rakesh Rao Ramachandra Rao, Rasmus Merten, and Alexander Raake. Appeal and quality assessment for AI-generated images. pages 115–118, 06 2023.
- [71] Jianning Chi, Zhiyi Sun, Huan Wang, Pengfei Lyu, Xiaosheng Yu, and Chengdong Wu. CT image super-resolution reconstruction based on global hybrid attention. *Computers in Biology and Medicine*, 150:106112, 2022.
- [72] Divya Saxena and Jiannong Cao. Generative adversarial networks (GANs survey): Challenges, solutions, and future directions, 2023.
- [73] Yoonsuh Jung and Jianhua Hu. A K-fold averaging cross-validation procedure. *Journal of Nonparametric Statistics*, 27(2):167–179, 2015.
- [74] Na Sun and Huina Li. Super resolution reconstruction of images based on interpolation and full convolutional neural network and application in medical fields. *IEEE Access*, 7:186470–186479, 2019.
- [75] Satoru Mizusawa, Yuichi Sei, Ryohei Orihara, and Akihiko Ohsuga. Computed tomography image reconstruction using stacked U-Net. *Computerized Medical Imaging and Graphics*, 90:101920, 2021.
- [76] Loris Nanni et al. Comparison of different convolutional neural network activation functions and methods for building ensembles for small to midsize medical data sets. *Sensors*, 22(16):6129, 2022.
- [77] Fabio Garcea, Alessio Serra, Fabrizio Lamberti, and Lia Morra. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*, 152:106391, 2023.
- [78] Mürsel Ozan İncetaş. Image interpolation based on spiking neural network model. *Applied Sciences*, 13(4), 2023.
- [79] Radiopaedia.org. Hounsfield unit. *Radiopaedia*, 2023.
- [80] W Pu, B Sober, N Daly, C Zhou, Z Sabetsarvestani, C Higgitt, I Daubechies, and MRD Rodrigues. Image separation with side information: A connected auto-encoders based approach. *IEEE Transactions on Image Processing*, 32:2931–2946, 2023.
- [81] You Sheng Toh and Carol Anne Hargreaves. Analysis of 2D and 3D convolution models for volumetric segmentation of the human hippocampus. *Big Data and Cognitive Computing*, 7(2), 2023.
- [82] Vivek Bakaraju and Kishore Reddy Konda. Only sparsity based loss function for learning representations. *arXiv preprint arXiv:1903.02893*, 2019.
- [83] AB Abdusalomov, R Nasimov, N Nasimova, B Muminov, and TK Whangbo. Evaluating synthetic medical images using artificial intelligence with the GAN algorithm. *Sensors (Basel)*, 23(7):3440, 2023.
- [84] Salih Saleh. Generating synthetic CT images using diffusion models, 2023.
- [85] Shaoyan Pan, Tonghe Wang, Richard L J Qiu, Marian Axente, Chih-Wei Chang, Junbo Peng, Ashish B Patel, Joseph Shelton, Sagar A Patel, Justin Roper, and Xiaofeng Yang. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine Biology*, 68(10):105004, 2023.